Borrower is requisted to check the book and get the signatures on the torned pages, if any.

THE ROYAL SIGNALS

HANDBOOK OF

LINE COMMUNICATION

A comprehensive text-book dealing with the theoretical and practical aspects of the transmission of intelligence over lines

VOLUME I

MCMXLVII

First Published 1947

Reprinted 1951 Incorporating Amendments 1, 2, 3 and Errata

PRINTED AND PUBLISHED BY HIS MAJESTY'S STATIONERY OFFICE

To be purchased from

York House, Kingsway, LONDON, W.C.2   429 Oxford Street, LONDON, W.1

P.O. Box 569, LONDON, S.E.1

13a Castle Street, EDINBURGH, 2   1 St. Andrew's Crescent, CARDIFF

39 King Street, MANCHESTER, 2   Tower Lane, BRISTOL, 1

2 Edmund Street, BIRMINGHAM, 3   80 Chichester Street, BELFAST

or from any Bookseller

Price £1 0s 0*d* net

PRINTED IN GREAT BRITAIN

S.O. Code No. 57–607*

FOREWORD

A modern text-book covering the theory and practice of line munications has long been required by the Army. This book been produced at the School of Signals; it is in two volumes, first volume dealing with the basic theory required for a study ne communications and including a summary of the mathematics essary for this, the second volume dealing with the practical plications to military line equipments.

The invaluable assistance of the Post Office Engineering Depart-ent in the preparation of this book is gratefully acknowledged.

Major-General,
Director of Signals.

The War Office,
January, 1947.

CONTENTS

VOLUME I

LINE COMMUNICATION

CHAPTER 1

AN INTRODUCTION TO LINE COMMUNICATION

Communication is the conveying of information from one place to another. This may be accomplished in many different ways; for example, by speaking or shouting, by sending a written message, or by flashing a lamp

To send information from one place to another, three things are necessary: a sending device, a receiving device, and some form of link between the two.

In order to be sent from the transmitter to the receiver, the information must be translated into "signals" that can be passed along the link concerned. In line communication, the signals must be in the form of electric currents that can be transmitted along electrical conductors (referred to as lines), and reconverted into an intelligible form at the receiver. It is the purpose of this book to study the behaviour of electric currents in such lines, and the principles of the various types of terminal equipment.

Line communication is divided into two classes, *viz.* "telegraphy" and "telephony", according to the nature of the signals; the general principles of these two classes will be briefly outlined in the ensuing sections of this chapter.

TELEGRAPHY

Telegraphy may be defined as the art of transmitting messages by means of "code-signals". Telegraph codes suitable for line working are usually built up from two basic conditions, called "mark" and "space". The code-signals representing the various letters of the alphabet, numerals, and punctuation signs, accordingly consist of different combinations of marking and spacing elements.

These signals may be sent by means of a hand-operated key, and received by some instrument for converting them into aural or visual signals that can be translated by a receiving operator; this is called "manual" telegraphy. In "semi-automatic" telegraphy, an operator is again required to manipulate the transmitting

instrument, but at the distant end the signals are received and recorded on paper by a machine that does not necessarily require the continuous presence of an operator. In fully automatic telegraph systems, the actual transmitting and receiving instruments operate without human assistance, the message being fed to the transmitter in the form of a paper tape that is suitably prepared at any convenient time prior to transmission.

Telegraph codes

There are two codes at present in common use for line telegraphy, the Morse code and the Murray code.

The Morse code is the standard code used for manual telegraphy; it is also extensively used for automatic wireless telegraphy. In it, the signals representing different letters, *etc.*, are, in general, of

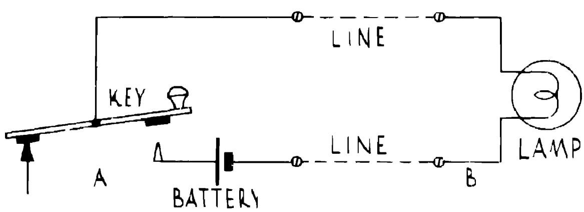


FIG. 1.—Basic telegraph circuit.

different duration, since they consist of different numbers of marking and spacing elements. The latter occur in "groups" as follows:—

(*a*) A single marking element, called a "dot".

(*b*) Three successive marking elements (with no interval between them), forming a long marking signal called a "dash".

(*c*) A single spacing element (equal in length to a dot), to separate the dots and dashes forming a character.

(*d*) Three successive spacing elements, to separate the letters of a word.

(*e*) Five successive spacing elements, to separate words.

The Murray code was especially designed for automatic working, and it differs basically from the Morse code in that every group in it consists of the same number (five) of elements, each of which may be either mark or space.

This code is used in the teleprinter, but in addition, special "start" and "stop" elements are used at the beginning and end of each group to separate the letters of a word. A special group has to be used to separate words.

Basic telegraph circuit

A simple telegraph circuit is shown in Fig. 1. A key and battery are connected to one end *A* of a line, and a lamp to the other end *B*. When the key is not pressed, the battery is disconnected from the

line, and the lamp will not light ; this is the " spacing condition ". Depression of the key will connect the battery to the line, and the lamp will light ; this is the " marking condition ". The Morse code may therefore be used to convey intelligence from *A* to *B*.

Practical developments

Since watching the lamp would be very tiring for the receiving operator, the lamp might be replaced by some device that would produce two different sounds, or alternatively sound and no sound, to represent the marking and spacing conditions. Alternatively, a permanent record of the received signals could be made by arranging an electro-magnet to deflect a pen over a moving paper tape. This would then leave a trace as in Fig. 2*a* if the pen were to be deflected sideways, or as in Fig. 2*b* if up-and-down. The latter form of trace gave rise to the terms " mark " and " space ".

For higher operating speeds an automatic or semi-automatic system must be used. An example of the latter is the teleprinter ; a " typewriter pattern " keyboard is provided, and when any key

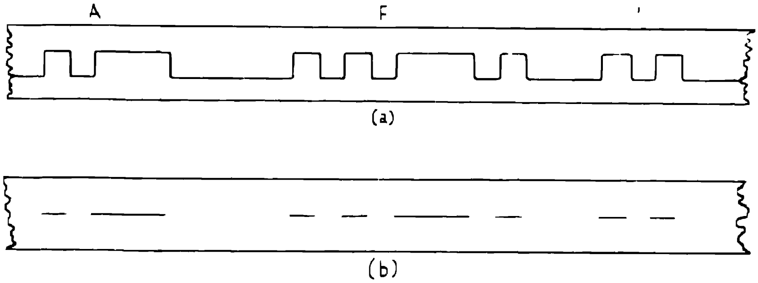


FIG. 2.—Recorded telegraph signals.

is depressed the appropriate groups of the Murray code are sent to line. The receiver converts the code-groups into mechanical movements that select and print the corresponding characters. By providing a transmitter and a receiver at each end, the simultaneous transmission of messages in the two directions can be obtained.

TELEPHONY

Telephony is the transmission of sound—in particular, speech—to a distant place. In the case of line telephony, the mechanical energy of the speaker's voice is made to control electric currents having similar characteristics. At the receiving station, these currents are reconverted into sound waves similar to those originated by the speaker.

Sound

Sound consists of air vibrations, which spread out in all directions from the source. The latter is usually a vibrating body—such as

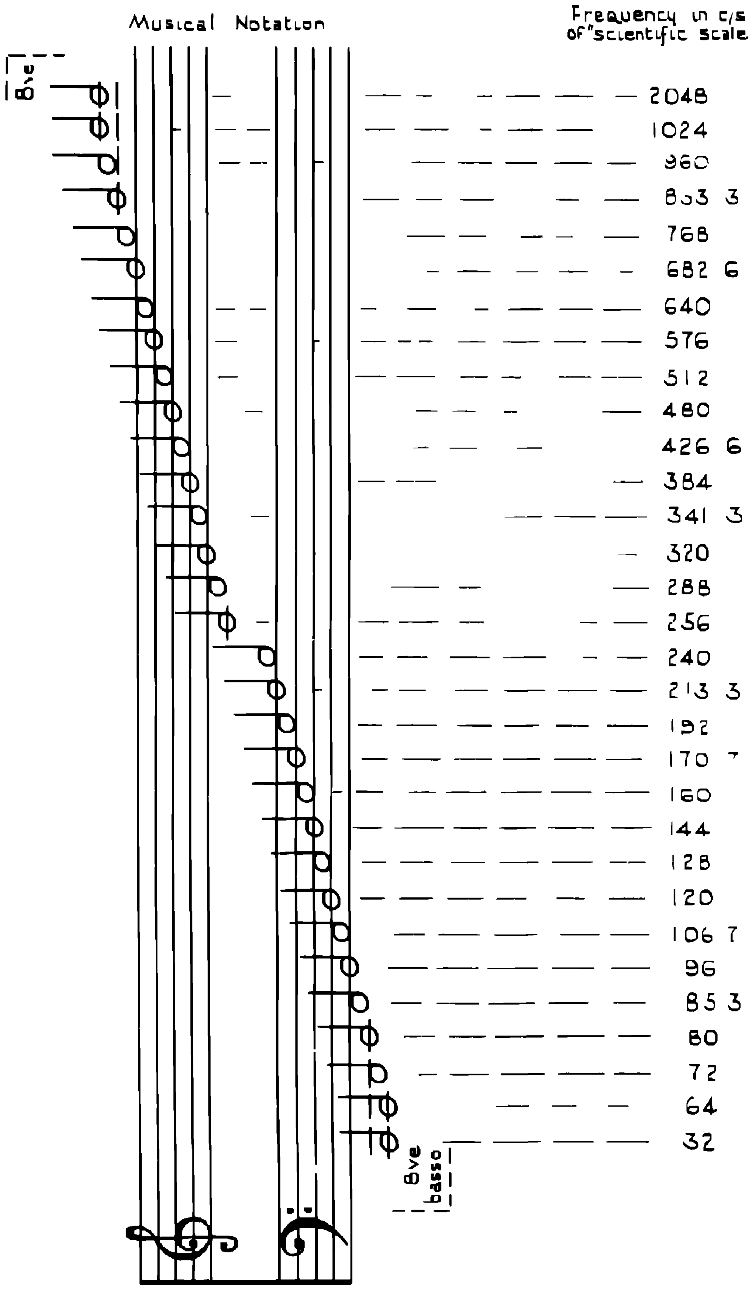


FIG 3 —Scientific musical scale.

the human vocal chords—which causes the adjacent air to vibrate in a like manner. The vibrations then spread out, at a speed of roughly 1,100 feet per second, causing all near-by air particles to oscillate longitudinally about their original positions. It must be noted that it is the vibrations, and not the air itself, which spread out from the source. When these vibrations reach any light free-to-move object, they will cause it to vibrate in a similar manner. If this object be the diaphragm of a person's ear, the vibration will give him the "sensation" of sound, the loudness depending on the amplitude of the vibrations.

The pitch of a sound depends on the "frequency" of the vibrations—that is to say, the number of vibrations per second; the greater the frequency, the higher the pitch. The ear can hear sounds of frequencies from about 20 to 20,000 cycles per second, the exact audible limits varying from person to person. However, if only the frequencies ranging from 300 to 2,000 c/s are received, the speech will still be intelligible although its quality will be changed. Fig. 3 shows the frequencies of the various notes of the "scientific" musical scale.

Waveform

The simplest form of vibration is the rather dull and insipid note produced by a tuning fork. This sound is called a "pure" tone, since it consists of a "sinusoidal" vibration of one frequency only, and can be represented by a "sine" curve, as in Fig. 4*a*. In general, however, the vibrations corresponding to the tones of musical instruments and speech are somewhat more complex (Fig. 4, *b* to *e*), but can be analysed into a "fundamental" vibration—of a frequency determining the pitch of the note—together with a number of higher-frequency "harmonics" that determine the quality or "timbre" of the sound. Each of these harmonics is a simple (sinusoidal) vibration of frequency equal to a multiple of the frequency of the fundamental, and it is the relative proportions of the various harmonics which distinguish between the tones of, say, a violin and a trumpet, both playing the same note of the scale, or between the various sounds that occur in speech, or between different people's voices. If a telephone circuit limits the transmitted frequency band to 300–2,000 c/s, the higher harmonics will be lost, and although speech will still be perfectly intelligible, this distinction will be lost, and it may be difficult to recognise the voice of the person speaking at the other end.

Basic telephone circuit

Fig. 5 shows a simple telephone system. The battery passes a current through the microphone, along the line, and through the receiver at the distant station. The microphone (see Fig. 6) is in effect a resistance that varies according to the position of its diaphragm. Sound vibrations falling on the diaphragm cause it to move forwards and backwards, and so cause corresponding changes

in resistance. The current in the line will therefore vary at the frequency of the sound vibration, and in a manner representing the timbre of the sound. The receiver (Fig. 7) consists of an iron diaphragm in front of an electro-magnet, towards which it is attracted; being held round the rim, it is drawn in at the centre

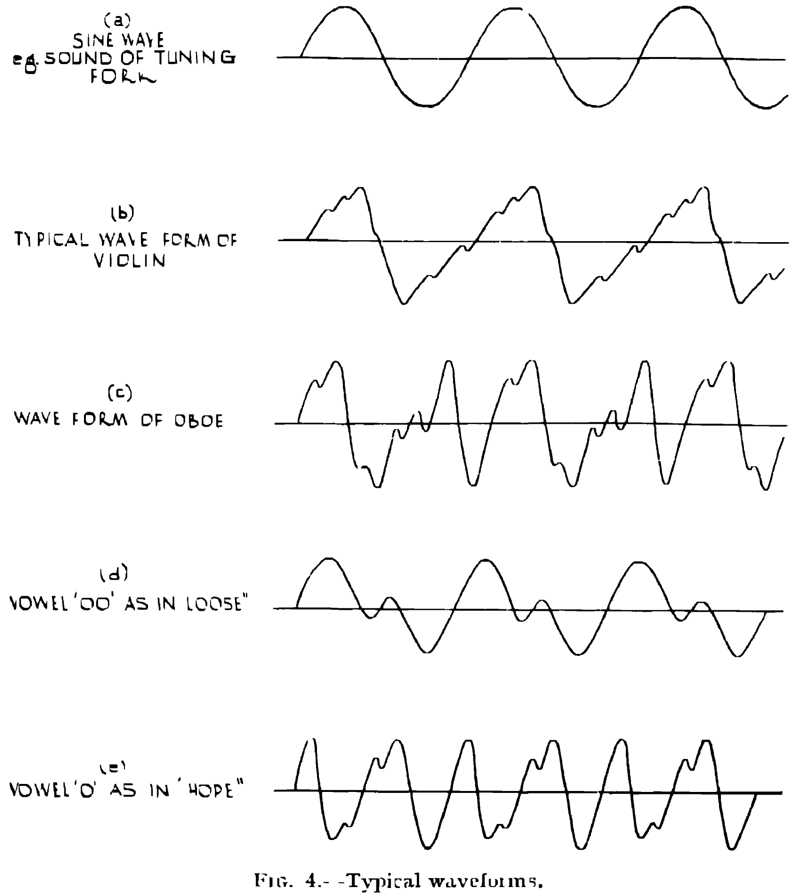


FIG. 4.—Typical waveforms.

towards the magnet. The strength of the attraction, and therefore the extent of the movement, depends on the current through the coils of the electro-magnet. Hence the diaphragm will move forwards and backwards in accordance with the current changes, and, in moving, will give rise to sound waves similar to those falling on the microphone at the far (transmitting) end of the line.

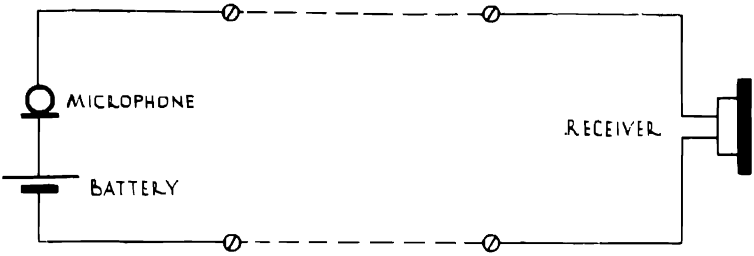


FIG. 5.—Simple telephone circuit.

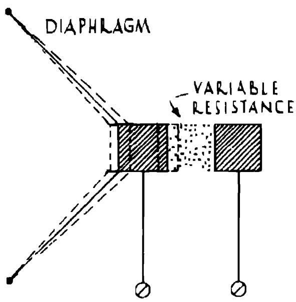


FIG. 6.—Microphone.

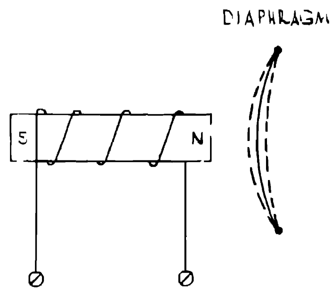


FIG. 7.—Telephone receiver.

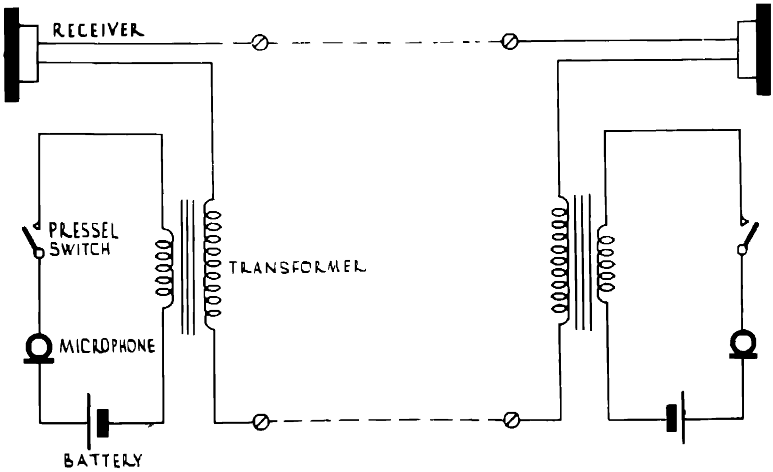


FIG. 8.—Basic telephone circuit.

Practical developments

While the circuit of Fig. 5 would be quite satisfactory for working over a short line, it would be useless on a long line with appreciable resistance, since the changes in resistance of the microphone would be negligible compared with the resistance of the line, and so the changes in current would be barely perceptible. The difficulty is overcome in practice by the use of a transformer, as in Fig. 8. This consists of two windings round a common core. The first is connected to the battery and microphone, and is of low resistance, so that the changes in resistance of the microphone can cause a large change in current round this circuit ; this winding is called the " primary ". The second winding consists of a large number of turns, so that the changes in primary current will induce into it the optimum voltage for driving the current through the line to operate the receiver at the distant end ; this is called the " secondary " winding. Over short distances, an instrument such as this will work satisfactorily with " earth return "—that is to say, with one of the two line wires replaced by a good connection to the

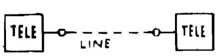


FIG. 9.—Two telephones connected by a line.

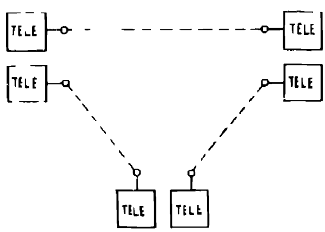


FIG. 10.—Three stations with independent telephones between each.

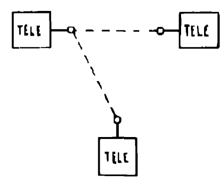


FIG. 11.—Three telephones with party line connection.

ground at each end. In addition to what is shown in the basic circuit of Fig. 8, telephones normally have some means of " calling ", or attracting the attention of the person at the other end, as, for example, a magneto generator, operation of which rings a bell on the distant instrument.

An important refinement usually incorporated is some form of " anti-sidetone " circuit, to prevent speech, and any other noises picked up by one's own microphone, from being heard in one's own receiver. This device is particularly advantageous when listening to weak signals in a noisy place.

EXCHANGES

The simplest possible telephone system consists of two telephones connected by a line as in Fig. 9. This is quite a suitable arrangement if two subscribers wish to speak to one another and to no one else. If three people wish to be interconnected, so that each can

converse privately with either of the other two, this can be arranged by repeating the simple lay-out, as shown in Fig. 10. This is uneconomical in equipment, however, since six telephones are needed to interconnect three subscribers.

If the subscribers are interconnected as shown in Fig. 11, only three telephones are used instead of six. This is not altogether satisfactory, however, because when any two of the three subscribers are conversing, the third can listen to the conversation; also, when one subscriber rings to attract the attention of another, the bell of the third subscriber rings too.

The switchboard

In practice, it is usually necessary for a subscriber to be able to speak privately with any one of a number of other subscribers. A flexible system is therefore needed whereby any subscriber's telephone can be connected at will to that of any other subscriber. The simple lay-outs described above are not suitable for this, and in practice this facility is provided by a "switchboard"; this is a piece of apparatus to which all the subscribers' instruments are connected (Fig. 12), and by means of which a switchboard operator can inter-connect any two subscribers. Many switchboards also provide for calls in which more than two subscribers are connected together.

By means of "junction" lines between switchboards, a subscriber connected to one switchboard can be connected through to a subscriber on another, as shown in Fig. 13. Moreover, two such junction lines can be connected together, as in Fig. 14; a subscriber on one switchboard can thus be connected "through" another switchboard, so that he can speak to any subscriber on any subsequent switchboard.

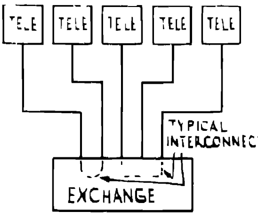


FIG. 12.—Telephones connected to exchange.

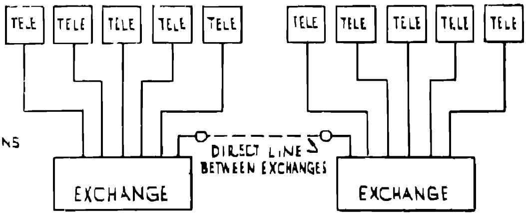


FIG. 13.—Two interconnected exchanges.

Larger switchboards (as distinct from portable field switchboards) generally have a certain amount of associated equipment separate from the switchboard itself. This includes such items as batteries, frames on which the lines are terminated, fuses and other protective devices, testing apparatus, *etc*. The term "exchange" is used to denote the switchboard together with all its associated equipment.

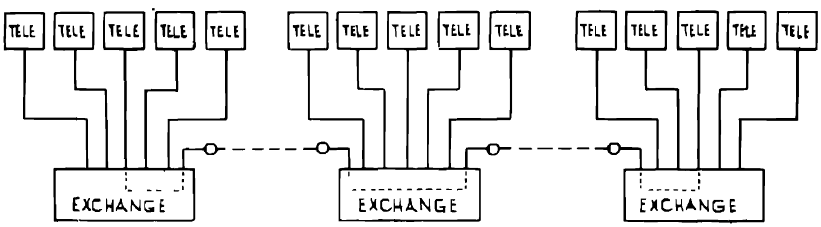


FIG. 14 Call through three exchanges.

Requirements of switchboards

(*a*) Subscribers must be able to attract the attention of the switchboard operator; one "calling indicator" is accordingly provided at the switchboard for each subscriber's line, and subscribers are provided with some means of sending a calling signal which will make the indicator operate.

(*b*) The switchboard operator must be able to attract the attention of any subscriber; each subscriber is therefore provided with a bell, or other form of alarm, that can be operated by a calling signal from the switchboard operator, and the switchboard must include a means of sending this signal.

(*c*) The switchboard operator must be able to connect any two or more subscribers' instruments together. This is sometimes done by keys on smaller switchboards but is more often effected by means of "cords", with plugs at the end, which the operator can insert into "jacks" connected to the various subscribers' lines.

(*d*) The subscribers must be able to inform the switchboard operator when they have finished their conversation, so that he can disconnect them. They do this by sending a "clearing signal", to operate some form of clearing or "supervisory" indicator at the switchboard.

(*e*) The switchboard operator must be able to speak to, and hear, any subscriber; he is therefore provided with a telephone that he can connect to any subscriber's line.

(*f*) The switchboard operator must be able to monitor, or "listen in to", any conversation without interfering with it, both to make sure that the call is really "through", and also to find out whether the subscribers have finished, since clearing signals may not always be given. He must therefore be able to connect his telephone to the subscribers' lines during a conversation.

(*g*) All the above requirements must also be satisfied when the "subscriber" is another exchange to which a direct line is provided.

Exchange signalling

From a technical point of view, the simplest method of calling a telephone exchange is the "magneto" system. In this system, the subscriber operates a magneto generator, which provides

sufficient power to operate the indicator at the switchboard. For signalling from subscriber to switchboard, this system suffers from the great disadvantage that if a subscriber fails to "ring off", there is nothing to tell the switchboard operator that the conversation has been concluded. Calling from the switchboard to the subscriber is, however, normally effected by the magneto system, since the ringing-off difficulty does not apply there, and a bell is the most satisfactory method of attracting a subscriber's attention. Another system sometimes used is the "buzzer"; in this system, a buzzer operated by the calling subscriber sends to line an alternating current that can operate an indicator at the exchange, or cause a buzz in the receiver at the distant end.

To avoid the need for ringing off, the "central battery signalling" (CBS) systems were developed. In these, raising the handset of the subscriber's telephone from its cradle or hook operates a switch and completes a DC circuit *via* the line, so that the battery at the exchange can operate the calling indicator. This is known as "loop" calling. The cord circuit includes a "supervisory" indicator, arranged to attract the operator's attention when the handset is replaced on its cradle, so that the subscriber does not have to ring off. Frequently, two supervisory indicators are provided, one connected to the line of each of the two subscribers conversing, so that independent supervision of each subscriber's actions is obtained; if one supervisory indicator operates, and not the other, then the operator knows that the one subscriber has replaced his handset, but that the other has not, and probably requires attention.

A diagram showing the basic circuits of various exchange systems will be found in Chapter 19, Vol. II.

The CB and automatic systems

In all the telephone systems considered above, batteries are required at each subscriber's instrument for the microphone circuit. In the case of large permanent installations this arrangement is undesirable, since it is necessary for a mechanic periodically to change the batteries at all the subscribers' telephones—which may be scattered over a large area. Several types of "common battery" (CB) systems have been developed, in which one large battery at the exchange is used to provide the energising current for all the subscribers' microphones; there are then no components in the subscribers' instruments that should, in the normal course of events, ever require attention. The lines between the subscribers and the exchange have to be of fairly low resistance. The system is therefore unsuitable for subscribers situated far from the exchange, and they will have to use CBS type instruments, which work quite satisfactorily on CB exchanges. In all CB systems, the exchange is called by means of a loop between the two legs of the line, as in CBS systems, and the subscriber is called by means of a generator and bell.

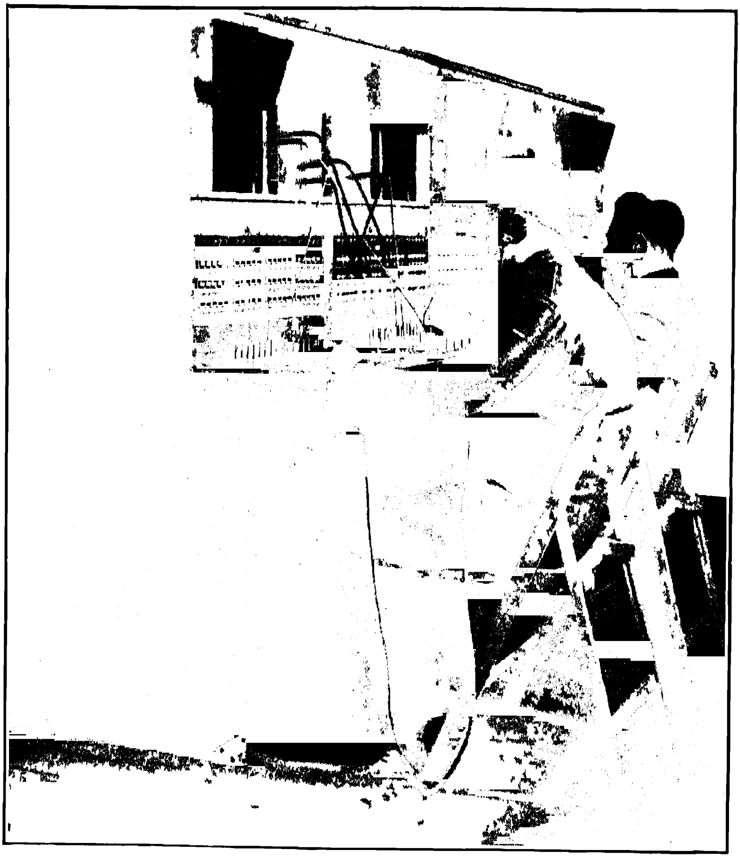


PLATE 1 Army magneto exchange

In automatic systems, raising of a subscriber's handset from its cradle loops the line, and causes him to be connected on to a "selector". "Dialling" any digit interrupts this loop a corresponding number of times, causing the selector to step up and round and to connect the subscriber to a further selector. This process is repeated for the second and subsequent digits, connecting the calling subscriber to the line of the subscriber whose number he has dialled. But before making this connection, the final selector connects the called subscriber's line to a power-driven ringing generator; when the called subscriber raises his handset from its cradle, his line is disconnected from the generator, and connected through to the calling subscriber. When the calling subscriber replaces his handset the selectors return to their normal position in readiness to deal with another call.

LONG LINES AND REPEATERS

Attenuation, distortion and interference

A line "attenuates" the signals passed along it; that is to say, the signals reaching the receiving end are weaker than those transmitted from the sending end, owing to power losses in the line itself. The attenuation increases as the length of the line is increased, and on long lines the received signals may be so weak as to be useless unless some steps are taken to make up for the attenuation by amplification; this is done by means of "repeaters". In addition to attenuation, distortion and interference must be considered.

"Distortion" is said to occur in a line or in any piece of equipment when the signal leaving it is not identical in "form" or character (though not necessarily in magnitude) with that entering it. The attenuation of a line increases with frequency, and consequently the higher frequency components of a signal are weakened more than the lower. Since the relative proportions of the various components of a signal determine its character, this variation of attenuation with frequency constitutes a form of distortion (attenuation-distortion) which, if not corrected, may render the received signal unintelligible. The required correction is effected by means of "equalisers".

Any line picks up noise from nearby power lines, *etc.*, and also "cross-talk" from neighbouring telephone and telegraph circuits. On a long line, the amplitude of such interference might be sufficient to "drown" the signals. Both forms of interference are in general much less in the case of metallic circuits than in the case of those using earth return, and can be greatly reduced by ensuring a symmetrical disposition of the two legs of the line.

Repeaters

To compensate for the power loss in a long line, "repeaters" can be employed to "boost up" the weakened signal currents.

Normally, two amplifiers will be required at each repeater, one for each direction of transmission; these are usually called "send" and "receive", or the "up to-down" and "down-to-up", amplifiers, to distinguish between them. Also, some form of "separating device" will be required between the telephone and the amplifiers to isolate the input circuit of the one amplifier from the output of the other, as in Fig. 15. The commonest form of separating device is the "hybrid transformer". If this isolation is not effected on both sides of the amplifiers, then a signal from one of the telephones, after amplification by the "send" amplifier, will reach the input side of the "receive" amplifier; after further amplification by this, it will get back to the input of the send amplifier, and the amplifiers will "sing" or "howl".

Repeaters must be spaced at intervals along the line. If one large amplifier were to be used at the transmitting end of the line,

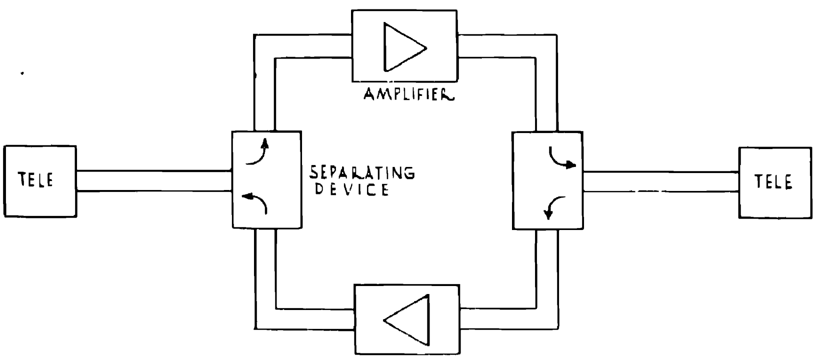


FIG. 15.—Telephone repeater.

its output would be so great that (*a*) it would cause severe interference to other circuits, and (*b*) the cost of the amplifier would be enormous. On the other hand, if all the amplification were to be concentrated at the receiving end of the line, the signal currents reaching the amplifier would, in the case of long lines, be much smaller than the noise and cross-talk currents picked up by the line; these would therefore obliterate the signal. By careful repeater spacing, however, the effects of line attenuation can be satisfactorily overcome, while the effects of noise and cross-talk can usually be reduced to negligible proportions.

Two- and four-wire circuits

It will be seen from Fig. 15 that whereas there are only two wires from the telephone, there are four wires between the amplifiers and the "separating device". On a long line requiring a number of repeaters, there are therefore two possible methods of working:

Two-wire Circuit.—Two wires can be run between repeaters, each of which will then have to contain two "separating devices", one on each side of the amplifiers, as in Fig. 16.

Four-wire Circuit.—By running four wires between repeaters, as in Fig. 17, only the two terminal repeaters need contain any separating devices; intermediate repeaters will each have two amplifiers, one for transmission in each direction, and these two amplifiers will be completely separate electrically, since different pairs of wires are used for the two directions.

Whereas the two-wire circuit is more economical in lines, a four-wire circuit requires, in general, fewer repeaters for a given

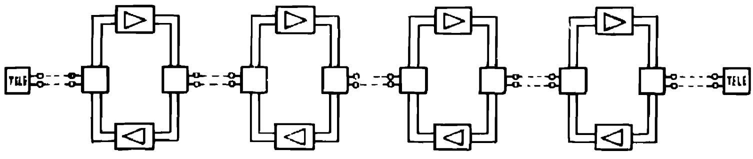


FIG. 16.—Two-wire repeatered circuit.

length of line, is easier to set up and maintain, and is more stable under varying line conditions. Electrically, the four-wire circuit is therefore preferable, particularly on long lines, but the economy in lines afforded by two-wire repeatered circuits is such that these are frequently used for military purposes.

Signalling over repeatered circuits

Speech amplifiers cannot deal with the low-frequency calling signals from a magneto generator, and repeaters must incorporate additional apparatus when generator signalling is used. Alternatively, the generator signals may be converted at the terminals into

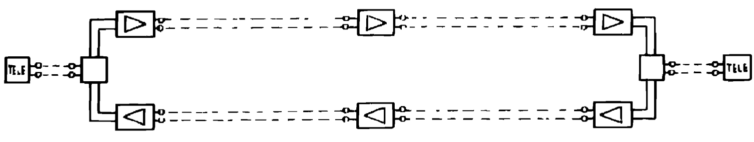


FIG. 17.—Four wire repeatered circuit.

tones lying within the speech range, which can then be amplified by the same apparatus as the speech. This latter method is known as "voice-frequency" signalling.

Equalisers

Equalisers are electrical networks designed to counteract the attenuation distortion occurring in any part of a circuit. They do this, either by introducing additional attenuation at those frequencies at which the attenuation is least, so that the overall attenuation is independent of frequency; or else by increasing the amplification in a repeater at those frequencies at which the attenuation is

greatest. Equalisers are usually included in repeaters, as well as in terminal equipment, so that the distortion, in the same way as the attenuation, can be corrected all along the line, before it can reach too serious a value.

SIMULTANEOUS TRANSMISSION OF SEVERAL MESSAGES OVER ONE LINE

Particularly in rear areas, very large numbers of telephone conversations are frequently required between different HQs, and a single line between the respective exchanges, such as that shown in Fig. 13, would therefore be totally inadequate. Owing to the time and the volume of stores required to install several lines, special forms of apparatus are often used that enable a line to handle more than one telephone or telegraph message at a time. This is particularly desirable in the case of long repeatered lines.

Superposing

"Superposing" is the direct connection of more than one telephone or telegraph instrument on to one line, and there are several simple methods of doing this. The simplest method is direct "series" superposition of a telegraph on to a telephone circuit, as in Fig. 18. This method is very restricted in its application. It depends on the principle that the telephone is unaffected by the DC and very low-frequency AC used by the telegraph

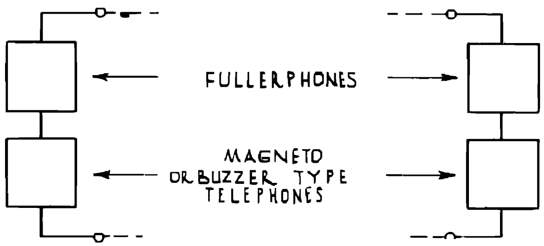


FIG. 18—Series superposing.

instrument, and that the telegraph instrument is not affected by the alternating currents of the frequencies used by the telephone. This arrangement cannot be used when working to an exchange employing direct current in the line (*i.e.* a CB or CBS exchange). The telephone instrument used must be capable of passing the direct current required by the telegraph instrument.

The only telegraph instrument suitable is the Fullerphone. Even using this instrument, there will be a certain amount of interference to the Fullerphone each time the buzzer or ringing magneto of the telephone is operated.

Phantom working

Phantom working is a more satisfactory method of superposing. In earth-return phantom working, one telephone circuit is obtained in the usual way between the two wires of the line, so that current goes out on one wire and returns by the other. The second instrument is connected between the centre-point of this "physical"

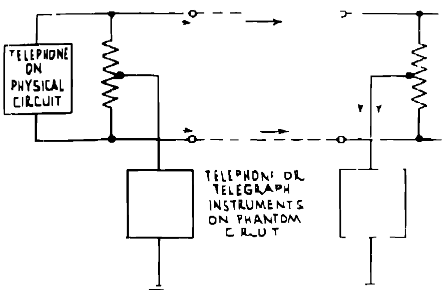


FIG. 19.—Earth-return phantom superposing.

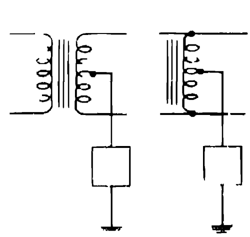


FIG. 20.—Phantom connections.

circuit and earth, as in Fig. 19, so that the current divides equally between the two wires—going out on both wires in parallel, and returning *via* the earth; this latter is called the "earth-return phantom" circuit. Fig. 20 shows two methods that can be used to obtain the centre tap on the physical circuit.

While there is usually no appreciable interference between the physical and the phantom circuits on a pair, the earth-return phantom circuit tends to pick up a lot of noise and interference, since it uses an earth return. On long lines, earth-return phantom circuits are not satisfactory as high grade speech circuits. They are, however, frequently used in forward areas, and are sometimes

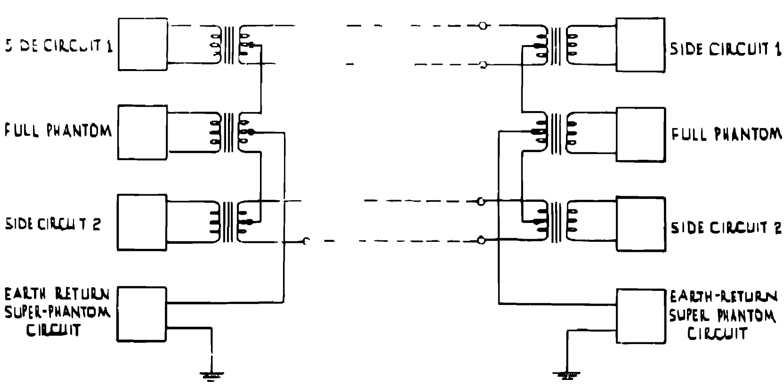


FIG. 21.—Full phantom working, with earth-return superphantom circuit.

used for " maintenance telephones " between repeater stations, to provide communication between the station mechanics.

An improvement on the earth-return phantom circuit just described is the " full phantom " circuit, in which the " phantom " telephone is connected between the centre-taps of two physical circuits, as in Fig. 21. An " earth-return super-phantom " circuit (sometimes known as double-phantom or ghost circuit) can then be added between the centre-tap of the full-phantom circuit and earth. The full-phantom circuit avoids interference from earth currents, but is suitable for high-grade speech circuits only when the two pairs on which it is superposed are of low capacity and are balanced with respect to each other as well as to earth. It is unsuitable for use on star quad cable, which has a relatively high capacity, but it may be used on multiple twin cables and " square-formation " permanent (overhead) line. Nevertheless, a high-grade circuit is obtained from a full phantom only when the physical pairs are very carefully balanced; the earth-return super-phantom circuit is clearly just as subject to interference from earth currents and cross-talk as the simple earth-return phantom.

V.F. telegraphy

The transmitted signals of the telegraph systems so far described have too low a frequency to be amplified conveniently by a valve amplifier in the same way that speech is amplified by a repeater. If, however, the telegraph signals are converted into " tone " within the speech range, they can then be sent over an ordinary telephone circuit and through the speech amplifiers. This method of transmitting telegraph signals is called " voice frequency " (VF) telegraphy.

It is possible to send these " voice frequency tone " signals over the same line as a telephone conversation, and at the receiving end to separate the VF tone from the speech so that there is no mutual interference. The use of this principle will thus permit one speech and one telegraph circuit to work simultaneously over one pair of lines. Although the advantages of this over phantom working are not at first apparent, it may be said that satisfactory operation of the telegraph equipment is possible over much greater distances. Also, in the case of a line sufficiently long to require a repeater for the telephone circuit, the ordinary telephone repeater will amplify the telegraph tone as well as the speech, with no modification at all, whereas with phantom working, a separate repeater would be necessary for the telegraph signals.

Alternatively, it is possible to send a number of different VF tones over the line instead of speech, thus forming a multi-channel VF telegraph system. As many as eighteen such VF channels may be employed over a telephone circuit (assuming, of course, the appropriate equipment at each end), and the advantages of the use of VF telegraphy then become apparent. In rear areas, where there is a vast amount of administrative traffic to be passed, one telephone

can be replaced by a number of telegraph instruments (*e.g.* teleprinters).

Carrier telephony

In carrier telephony, the signals from the transmitting telephone are altered in such a way that they can be passed over the same line as signals from an ordinary telephone, separated from these at the receiving end, and then altered back into normal type telephone currents similar to those leaving the transmitting instrument. This " alteration " consists in moving bodily the whole frequency band occupied by the speech signals to a higher position in the frequency spectrum (say, 3,000 to 6,000 c/s), so that whereas a line used for normal telephony has to pass only frequencies between 0 and 3,000 c/s, a line carrying one " physical " and one " carrier " telephone channel, would have to pass all frequencies between 0 and 6,000 c/s. Carrier telephone systems are therefore more stringent in their line requirements than ordinary telephone circuits.

In multi-channel carrier systems, the signals from each channel transmitter are raised to successively higher frequency bands, and consequently the lines for use with such equipment must be even more carefully chosen. This will be realised when it is stated that under certain conditions, a four-channel carrier telephone system may transmit frequencies up to 50,000 c/s.

Clearly, since all the channels of a multi-channel carrier telephone system are suitable for telephony, they are also suitable for VF telegraphy, and it is common practice to operate a six-channel VF telegraph system over one channel of a four-channel carrier telephone system; this then provides three telephone and six teleprinter circuits.

CHAPTER 2

MATHEMATICS

In order to understand many of the principles of line communication, some knowledge of mathematics is essential. This chapter contains a summary of the simpler branches of mathematics with which the reader is expected to have some acquaintance. The chapter is divided into two parts. Part I contains a summary of algebra and trigonometry, leading up to vectors and complex numbers. Part II includes the elements of calculus and hyperbolic functions, and is intended for those who wish to make a more detailed study of the subject.

Tables of logarithms, trigonometrical functions, *etc.*, will be found at the end of this volume (page 775).

Mathematical symbols

$=$	is equal to . . .
$\equiv$	is equivalent to . . .
$\simeq$	is approximately equal to . . .
$\neq$	is not equal to . . .
$\rightarrow$	tends to *or* approaches . . .
$>$	is greater than . . .
$\geqslant$	is greater than or equal to . . .
$\gg$	is very much greater than . . .
$\not>$	is not greater than . . .
$<$	is less than . . .
$\leqslant$	is less than or equal to . . .
$\ll$	is very much less than . . .
$\not<$	is not less than . . .
$\sim$	the positive difference between . . .
$\sqrt{}$	the square root of . . .
$\sqrt[n]{}$	the *n*th root of . . .
$\angle$	a positive angle; that is, an angle measured in an anti-clockwise direction.
$\overline{\diagdown}$	a negative angle; that is, an angle measured in a clockwise direction.

PART I

ELEMENTARY MATHEMATICS

Areas, surfaces and volumes

	Surfaces	*Volumes*
Sphere	$4\,\pi.r^2$	$\frac{4}{3}.\pi.r^3$
Cone	$\pi\, r.l + \pi r^2$	$\frac{1}{3}.\pi\, r^2.h$
Cylinder	$2\pi rh + 2\pi r^2$	$\pi r^2 h$
Triangular prism	$h(a + b + c) + 2\triangle$	$\triangle\,.\,h$
Cuboid	$2ab + 2bc + 2ca$	$a\,.\,b\,.\,c$
	Areas	
Circle	πr^2	
Sector	$\frac{1}{2}r^2\theta = \frac{\pi}{360}r^2\,.\,\varphi^\circ$	
Parallelogram	$hb = b.c.\sin A$	
Trapezium	$\frac{1}{2}h(x + y)$	
Triangle	$\frac{1}{2}hb = \frac{1}{2}b.c.\sin A = \sqrt{s(s-a)(s-b)(s-c)}$	

where r = radius
l = slant height
h = perpendicular distance between faces
x, y, a, b, c = sides
s = semi-perimeter = $\frac{1}{2}(a + b + c)$
θ = angle in radians
φ = angle in degrees
$\triangle$ = area of triangle

Pythagoras' Theorem

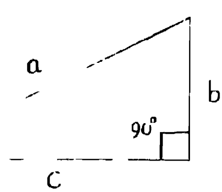


FIG. 22.—Pythagoras' Theorem.

In any right-angled triangle:—

$$a^2 = b^2 + c^2 \qquad (1)$$

where a is the length of the hypotenuse (*i.e.* the side opposite the right-angle), and b and c are the lengths of the other two sides.

Indices

a raised to the power n, written as "a^n", is defined as a multiplied by itself n times,

i.e. $\quad a^n = a \times a \times a \times a \ldots$ to n factors $\quad$ (2)

Thus $\quad 2^4 = 2 \times 2 \times 2 \times 2 = 16$

In such cases, n and 4 are known as indices.

Multiplication.—$a^n \times a^m = a^{n+m}$, *i.e.*, to multiply powers of the same number, the indices are added.

From this, it can be seen that $a^0 = 1$, for $a^n \times a^0 = a^{n+0} = a^n$, *i.e.*, multiplying by a^0 leaves the value unchanged,

$\therefore$ a^0 must be equal to 1

$(a^n)^m = a^{nm}$, *i.e.*, when a power is itself raised to a power, the two indices are multiplied.

$(ab)^n = a^n b^n$, *i.e.*, when the product of two numbers is raised to a power, the result is equal to the product of the two numbers each raised to that power.

$$\left.\begin{aligned} a^{-n} &= \frac{1}{a^n} \text{ (for } a^{-n} \times a^n = a^{-n+n} = a^0 = 1) \\ a^{\frac{1}{n}} &= \sqrt[n]{a}, \text{ for } \left(a^{\frac{1}{n}}\right)^n = a^{\frac{n}{n}} = a \end{aligned}\right\} \quad (3)$$

Note in particular that $\sqrt{ab} = \sqrt{a} \times \sqrt{b}$, and that $\sqrt{\dfrac{a}{b}} = \dfrac{\sqrt{a}}{\sqrt{b}}$

Surds.—A square root that cannot be further reduced is called a surd, *e.g.*, $\sqrt{2}$, $\sqrt{3}$, $\sqrt{5}$.

$\sqrt{8}$ is not a surd as it can be reduced to $2 \times \sqrt{2}$.

Rationalization

Evaluation of expressions such as $\dfrac{1}{\sqrt{2}}$ is difficult, since it involves division by a decimal. If top and bottom are both multiplied by $\sqrt{2}$ the value will be unaltered,

$$\therefore \frac{1}{\sqrt{2}} = \frac{1}{\sqrt{2}} \times \frac{\sqrt{2}}{\sqrt{2}} = \frac{\sqrt{2}}{2} = \frac{1.414}{2} = 0{\cdot}707,$$

and calculation is thus simplified.

Similarly, $\dfrac{1}{\sqrt{3} - \sqrt{2}}$ can be simplified by turning the bottom line into the difference of two squares —

$$\frac{1}{\sqrt{3} - \sqrt{2}} = \frac{1}{\sqrt{3} - \sqrt{2}} \times \frac{\sqrt{3} + \sqrt{2}}{\sqrt{3} + \sqrt{2}} = \frac{\sqrt{3} + \sqrt{2}}{(\sqrt{3})^2 - (\sqrt{2})^2}$$

$$= \frac{\sqrt{3}+\sqrt{2}}{3-2} = \sqrt{3}+\sqrt{2}$$

$$= 1 \cdot 732 + 1 \cdot 414 = 3 \cdot 146$$

Example.—

$$\frac{1}{\sqrt{1+x^2}+x} = \frac{1}{\sqrt{1+x^2}+x} \times \frac{\sqrt{1+x^2}-x}{\sqrt{1+x^2}-x}$$

$$= \sqrt{1+x^2}-x \quad \textit{Ans.}$$

Useful factorisations

$a^2 - b^2$	$= (a+b)(a-b)$	(4)
$a^3 - b^3$	$= (a-b)(a^2+ab+b^2)$	(5)
$a^3 + b^3$	$= (a+b)(a^2-ab+b^2)$	(6)
$a^4 + a^2b^2 + b^4$	$= (a^2-ab+b^2)(a^2+ab+b^2)$	(7)

Simultaneous equations

If there are n unknowns to be evaluated, there must be n equations. The procedure for solution is to reduce to $n-1$ equations with $(n-1)$ unknowns, and to continue this process until only one equation containing one unknown is left. The latter is a simple equation which can be solved at once; the other unknowns may then be found by working back through the equations.

Example.—

Solve

$$x+y-z=1 \quad \text{(i)}$$
$$2x+y+z=6 \quad \text{(ii)}$$
$$x+2y+2z=9 \quad \text{(iii)}$$

Adding equations (i) and (ii):—

$$x+y-z=1 \quad \text{(i)}$$
$$2x+y+z=6 \quad \text{(ii)}$$
$$3x+2y=7 \quad \text{(iv)}$$

Multiplying equation (i) by 2, and adding to equation (iii):—

$$2x+2y-2z=2 \quad \text{(i)} \times 2$$
$$x+2y+2z=9 \quad \text{(iii)}$$
$$3x+4y=11$$

Subtracting equation (iv) from (v):—

$$3x+4y=11 \quad \text{(v)}$$
$$3x+2y=7 \quad \text{(iv)}$$
$$2y=4 \qquad \therefore y=2$$

From (iv), $3x+4=7 \qquad \therefore x=1$

From (i) $1+2-z=1 \qquad \therefore z=2$

Thus:—

$$\left.\begin{array}{l} x=1 \\ y=2 \\ z=2 \end{array}\right\} \textit{Ans.}$$

Quadratic equations

Quadratic equations, *i.e.*, those of the form $ax^2 + bx + c = 0$, may be solved by three methods :—

(*a*) Factorisation.
(*b*) Completing the square.
(*c*) Formula.

(*a*) ***Factorisation.***—If the equation can be factorised by inspection as $(lx + m) . (px + q) = 0$, the roots are $x = -\frac{m}{l}$ and $x = -\frac{q}{p}$.

Example.—

Solve $3x^2 + 7x - 6 = 0$

$\therefore \ (3x - 2)(x + 3) = 0$

$\therefore \quad x = \frac{2}{3}$ or -3 *Ans.*

(*b*) ***Completing the square.***—Rearrange the equation to give a perfect square in terms of x, and take the positive and negative roots.

Example.—

$$x^2 + 4x + 2 = 0$$

Divide if necessary to make the coefficient of x^2 equal to 1, and then rearrange so that only terms containing x^2 and x appear on the left-hand side, *e.g.*, $x^2 + 4x = -2$.

Add (half the coefficient of $x)^2$ to each side :—

$$x^2 + 4x + 4 = -2 + 4$$

$$\therefore \ x^2 + 4x + 4 = 2$$

$$\therefore \quad (x + 2)^2 = 2$$

$$\therefore \quad x + 2 = \pm\sqrt{2}$$

i.e. $\quad x = -2 \pm \sqrt{2}$ *Ans.*

(*c*) ***Formula.***—Apply method (*b*) to the general equation :—

$$ax^2 + bx + c = 0$$

$$\text{Then } x^2 + \frac{b}{a}x = -\frac{c}{a}$$

Add $\left(\frac{1}{2}\frac{b}{a}\right)^2$, *i.e.* $\frac{b^2}{4a^2}$ to each side :—

$$x^2 + \frac{b}{a}x + \frac{b^2}{4a^2} = \frac{b^2}{4a^2} - \frac{c}{a}$$

$$\therefore \ \left(x + \frac{b}{2a}\right)^2 = \frac{b^2 - 4ac}{4a^2}$$

$$x + \frac{b}{2a} = \frac{\pm\sqrt{b^2 - 4ac}}{2a}$$

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \qquad (8)$$

This formula gives the two roots of any quadratic equation. Note the three cases :—

$b^2 > 4ac$, both roots are real,

$b^2 = 4ac$, the roots are equal,

$b^2 < 4ac$, both roots are imaginary.

Logarithms

Definition.—If $a^x = N$, then $x = \log_a N$ (read as : log, to the base a, of N) ; *i.e.*, the log to base a of N is the power to which a must be raised to produce N.

$$\begin{aligned} &e.g., & 3 &= \log_2 8, \text{ since } 2^3 = 8 \\ &\text{similarly,} & \tfrac{1}{3} &= \log_8 2 \\ &\text{and} & 0{\cdot}3010 &= \log_{10} 2 \end{aligned}$$

Antilogs.—If $x = \log_a N$ (*i.e.*, $a^x = N$), then $N = \text{antilog}_a\, x$ (read as : antilog, to base a, of x).

Multiplication by logs.—By definition, if $a^x = N$ and $a^y = M$, then :—

$$NM = a^x \times a^y = a^{x+y},$$

i.e.,
$$\log_a (NM) = x + y = \log_a N + \log_a M. \tag{9}$$

Similarly
$$\log_a \frac{M}{N} = \log_a M - \log_a N. \tag{10}$$

Powers.—
$$\log_a M^n = n \log_a M \tag{11}$$

$$\log_a \frac{1}{M} = \log_a M^{-1} = -\log_a M \tag{12}$$

Note that the log of 1, to *any* base, is equal to 0.

Change of base.—

$$\text{Let } \log_a N = x$$

$$\therefore \quad N = a^x$$

Taking logarithms, to base b, of each side :—

$$\log_b N = x \times \log_b a$$

$$\therefore \quad \log_b N = \log_a N \times \log_b a$$

i.e.
$$\log_a N = \frac{\log_b N}{\log_b a} \tag{13}$$

Thus the logarithm of a number to base a is equal to the logarithm of the same number to base b, divided by the logarithm of a to base b.

Use of logarithms.—If logarithms of numbers can be found, their use simplifies the processes of multiplication, division, *etc.* For if two numbers M and N have to be multiplied, their logs have to be added, which is a simpler process. For numerical calculations it is most convenient to take 10 as the base. The log of any number

between 1 and 10 will lie between 0 and 1, and the logs of such numbers are given in tables, usually to four figures. The log of *any* number can be found from these.

Thus $\log_{10} 2 = 0\cdot3010$

$$\log_{10} 20 = \log_{10}(2 \times 10) = \log_{10} 2 + \log_{10} 10$$
$$= 0\cdot3010 + 1 = 1\cdot3010$$

Similarly, $\log_{10} 2000 = 3\cdot3010$.

The number before the decimal point is called the "characteristic": the decimal is called the "mantissa".

The rule for calculating the log, to the base 10, of any number is to find from the tables the log to base 10 of the significant figures; this will give the mantissa. The characteristic is then written down —it will be equal to one less than the number of figures before the decimal point.

Example.– To find $\log_{10} 371\cdot9$.

Look up $\log_{10} 3719$. The tables give the mantissa as $0\cdot5704$.

The number of figures before the decimal point is 3

$\therefore$ the characteristic $= 2$

$\therefore \log_{10} 371\cdot9 = 2\cdot5704$. *Ans.*

Numbers less than 1.—In this case the log will be negative; to simplify calculation, however, the mantissa is kept positive, and the characteristic is made negative, and numerically equal to one more than the number of noughts after the decimal point.

Thus $\log_{10} 0\cdot03719 = \log_{10} \dfrac{3\cdot719}{100} = 0\cdot5704 - 2$, which is written $\bar{2}\cdot5704$, the bar over the "2" indicating that this digit is negative.

Multiplication and division.

To multiply, the logs are added.

To divide, the logs are subtracted.

Example.— Evaluate $\dfrac{13\cdot25 \times 0\cdot00137}{0\cdot1925}$

From the tables:	$\log_{10} 13\cdot25$	$1\cdot1222$
and:	$\log_{10} 0\cdot00137$	$= \bar{3}\cdot1367$
On adding:	$\log_{10}(13\cdot25 \times 0\cdot00137)$	$= \bar{2}\cdot2589$
From the tables:	$\log_{10} 0\cdot1925$	$= \bar{1}\cdot2844$
On subtracting:	$\log_{10} \dfrac{13\cdot25 \times 0\cdot00137}{0\cdot1925}$	$= \bar{2}\cdot9745$

$$\frac{13\cdot25 \times 0\cdot00137}{0\cdot1925} = \text{antilog}_{10}\,\bar{2}\cdot9745$$

From the tables, $\text{antilog}_{10}\,\bar{2}\cdot9745 = 0\cdot0943$. *Ans.*

Indices.—To obtain any power of a number, its log must be multiplied by the power.

Example.— Evaluate $(2\cdot38)^7$ by logs.

$$\log_{10} 2\cdot38 = 0\cdot3766$$
$$\text{Multiplying by 7: } \quad 7$$
$$\log_{10}(2\cdot38)^7 = \bar{2}\cdot6362$$
$$\therefore \quad (2\cdot38)^7 = \text{antilog}_{10}\, 2\cdot6362 = 432\cdot7 \quad \textit{Ans.}$$

Roots are similarly found.

Example.— Evaluate $\sqrt[4]{0\cdot0892}$ by logs.

$$\sqrt[4]{0\cdot0892} = (0\cdot0892)^{\frac{1}{4}}.$$
$$\log 0\cdot0892 = \bar{2}\cdot9504.$$

Some way must be found of dividing the $\bar{2}$ by 4. This is done by so expressing it, that the negative portion is a multiple of 4. Thus $\frac{1}{4} \times \bar{2}\cdot9504 = \frac{1}{4} \times (-4 + 2\cdot9504) = -1 + 0\cdot7376 = \bar{1}\cdot7376$

$$\sqrt[4]{0\cdot0892} = \text{antilog}\ \bar{1}\cdot7376 = 0\cdot5466. \quad \textit{Ans.}$$

Summations

The sign Σ (Greek capital letter sigma) is used to signify summation, and the limits (if any) over which summation is to be effected are indicated above and below it.

For example :—

$$\sum_{r=1}^{r=n} r^2 = 1^2 + 2^2 + 3^2 + \ldots\ldots + n^2$$

$$\sum_{x=3}^{x=5} x^3 = 3^3 + 4^3 + 5^3 = 216$$

Arithmetical progression. e.g., 1, 4, 7, 10, 13, ...

General form :—a, $(a + d)$, $(a + 2d)$, . . .

The r^{th} term is $a + (r - 1)d$

Sum of n terms is $\displaystyle\sum_{r=1}^{r=n} [a + (r - 1)\,d] = \tfrac{1}{2}n\,[2a + (n - 1)\,d]$

$$= \tfrac{1}{2} \times (\text{number of terms}) \times (\text{sum of first and last terms}). \quad (14)$$

Geometrical progression.—e.g., 1, 3, 9, 27, 81, ...

General form :—a, ax, ax^2, . . . The r^{th} term is ax^{r-1}

Sum of n terms is :— $\displaystyle\sum_{r=1}^{r=n} ax^{r-1} = \frac{a\,(x^n - 1)}{x - 1}$ (15)

If the numerical value of x is less than 1 (written $|x| < 1$), this sum tends to $\dfrac{a}{1 - x}$ as the number of terms increases.

Thus the sum of this series to infinity is given by :—

$$\sum_{r=1}^{r=\infty} ax^{r-1} = \frac{a}{1 - x} \quad (16)$$

Miscellaneous summations.

$$\sum_{r=1}^{r=n} r^2 = \frac{n}{6}(n+1)(2n+1) \qquad (17)$$

$$\sum_{r=1}^{r=n} r^3 = \left[\frac{n}{2}(n+1)\right]^2 \qquad (18)$$

$$\sum_{r=1}^{r=n} r(r+1) = \frac{n}{3}(n+1)(n+2) \qquad (19)$$

Means

The arithmetic mean of two quantities a and b is $\frac{a+b}{2}$.

If there are n quantities $a, b, c, d, \ldots$, their arithmetic mean is :—

$$\frac{a+b+c+d+\ldots}{n}$$

The geometric mean of two quantities a and b is $\sqrt{a \cdot b}$, i.e., the square root of their product.

Functional notation

Expressions such as $x^2 + 2x + 3$, $x + \frac{1}{x}$, $\sqrt{x^2 - 2}$ are expressions involving x, that is, their value depends on the value of x. Such expressions are known as "functions" of x.

It is useful to have a general notation for any function of x, hence a function of x is written as $f(x)$ or $F(x)$ or $\phi(x)$.

Thus the equation :—

$$y = x^2 + 2x + 3$$

becomes : $y = f(x)$

where : $f(x)$ stands for $x^2 + 2x + 3$.

This notation is extended to give the value of the function when x assumes some particular value, e.g., $f(2)$ is the value of $f(x)$ when $x = 2$.

If :— $f(x) = x^2 + 2x + 3$

then :— $f(2) = 2^2 + 2 \cdot 2 + 3 = 11.$

similarly : $f(3) = 3^2 + 2 \cdot 3 + 3 = 18.$

and :— $f(0) = 0 + 0 + 3 = 3$

GRAPHS

By drawing two axes, the position of any point in a plane may be fixed by its two "co-ordinates". These two axes are known as the x-axis (denoted by Ox), and the y-axis (denoted by Oy). The point O is known as the origin.

If PN is perpendicular to Ox, as shown in Fig. 23, then the "x-co-ordinate" of P is $ON = x$, and the y-co-ordinate is $NP = y$. The point P is referred to as the point (x, y). x is considered negative if P lies to the left of Oy; similarly, y is negative if P is below Ox. Thus to any point in the plane there corresponds one set of two numbers, and vice versa.

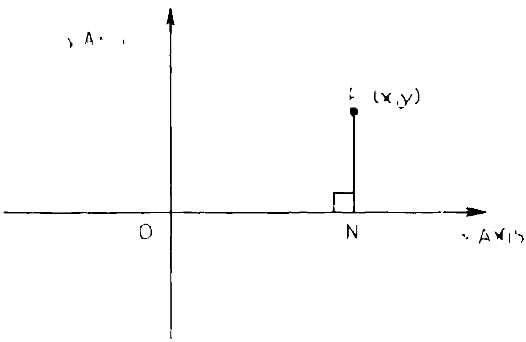


FIG. 23.—Use of co-ordinates x, y to determine position of a point P

If the values of x and y for the point P are unrestricted, P may be anywhere in the plane. If, however, the values are restricted by stating some definite relationship between them, P can lie only in certain positions. For a normal relationship between x and y (in the form of an equation) P will, in general, lie on a curve; and a curve of some sort exists for every equation. Much information can be obtained from a study of these curves or "graphs".

There are two types of equation: an "explicit" equation, such as $y = \dfrac{3x^2 + 5}{1 - x}$, where y is given at once as some "function" of x; and an "implicit" equation, such as $3x^2y + 5y^3 - 4x = 3$, where y is not given *directly* in terms of x.

Linear equations

An equation of the form $ax + by + c = 0$ always represents a straight line. To draw the line, it is best to find two points on it, and usually the easiest to find are those where $x = 0$ and where $y = 0$.

For example, consider the equation $2x - 3y = 6$.

If $x = 0$, $y = -2$ $\therefore$ $(0, -2)$ is on the line.

If $y = 0$, $x = 3$ $\therefore$ $(3, 0)$ is on the line.

The graph can now be drawn as a straight line through these two points (Fig 24).

The graph may be verified; for example, the point $(6, 2)$ should be on it, since $2 \times 6 - 3 \times 2 = 12 - 6 = 6$.

Solution of simultaneous equations.—Graphs may be used to solve simultaneous equations

$$\text{e.g.} \quad 2x - 3y = 6$$
$$\text{and} \quad x - y = 6.$$

Draw the graph of $x - y = 6$ on the same axes as the first, by drawing a straight line through the points (6, 0) and (0, − 6) (Fig. 24).

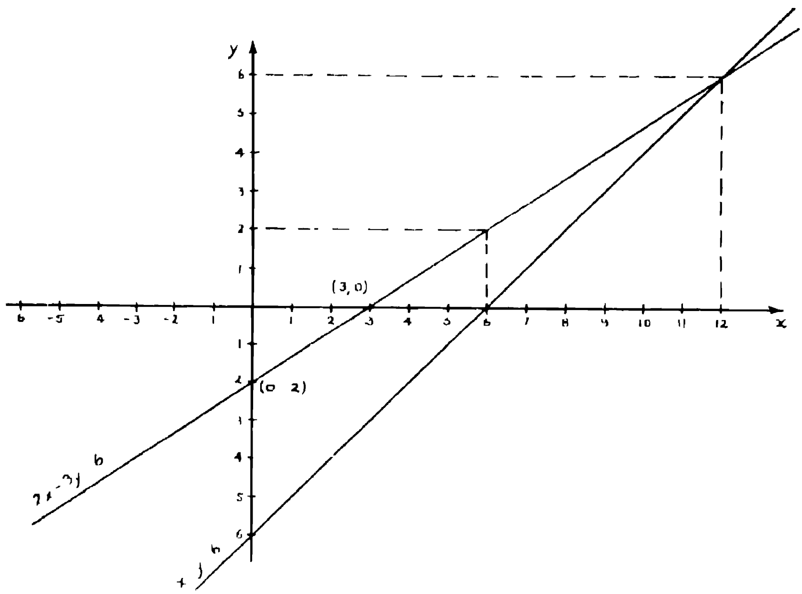


FIG. 24.—Graphs of $x - y = 6$ and $2x - 3y = 6$.

The point which represents the solution of the two equations must satisfy each equation, *i.e.* it must lie on each line. The only point which does this is, of course, the point of intersection—in this case (12, 6). Hence " $x = 12, y = 6$ " is the solution.

Other equations

In some cases, the form of the equation will indicate the shape of the curve. Consider, for example, the equation $x^2 + y^2 = a^2$.

This may be written as $\sqrt{x^2 + y^2} = a$.

But $\sqrt{x^2 + y^2}$ is the distance of the point P from the origin; therefore the equation represents a circle with centre at the origin and radius a.

Similarly $(x - l)^2 + (y - m)^2 = a^2$ is a circle with centre (l, m) and radius a.

Asymptotes

Many curves approach infinity along some straight line, as in Fig. 25.

The straight line is known as an "asymptote": note that a curve which goes to infinity along an asymptote must also return from infinity along it.

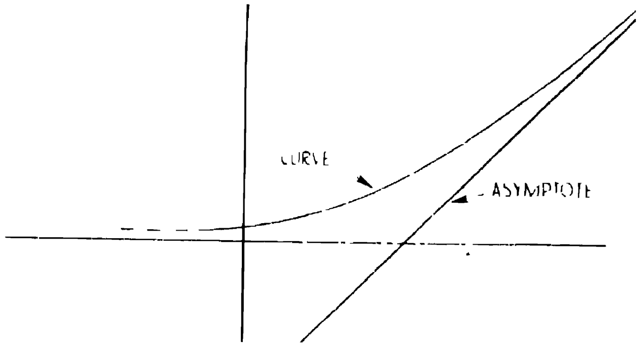


FIG. 25.—Asymptote.

To find the asymptotes of any curve, the condition which makes x or y infinite must first be found; an approximate form of the equation for large values of x or y can then be found.

General rules for plotting curves

1. Plot all points where x or y are zero or infinite.
2. Insert asymptotes, if any, finding out on which side the curve lies.
3. Plot any obviously important points.
4. Note any symmetry that may exist.
5. Avoid giving x or y various numerical values at random.

Example 1.—

Plot the graph of:— $y = x - \dfrac{1}{x}$

When $x = \pm 1$, $y = 0$, giving two points.

$x = 0$ makes y infinite, hence $x = 0$ is an asymptote. Find on which side of the asymptote the curve lies. If x is slightly greater than zero, y will be large and negative; hence the curve will go to infinity downwards along the right-hand side of the y axis. Similarly, if x is slightly less than zero, y will be large and positive, and the curve will therefore go up to infinity along the left-hand side of the y axis.

When $x \to \infty$, y also $\to \infty$, but if x is large $y \simeq x$, and hence $y = x$ is an asymptote. By the same method as in the last paragraph, it can be shown that the curve lies below the asymptote if x is positive and above it if x is negative.

Drawing in these points and asymptotes, Fig. 26*a* is obtained. It can easily be seen that the complete curve will be of the form

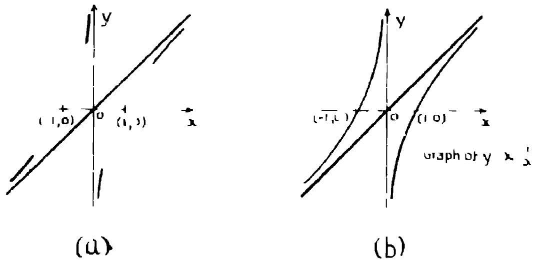


FIG. 26.—Graph of $y = x - \frac{1}{x}$

shown in Fig. 26*b*. Note that it is in two parts; this curve is known as a hyperbola.

Example 2.—

Plot the graph of: $$y = \frac{x-4}{(x-2)^2(x-3)}$$

When $x = 4$, $y = 0$, therefore $(4, 0)$ is on the curve. y also equals 0 if $x \to \pm\infty$; and, whether x is positive or negative, y will be positive for sufficiently large numerical values of x. Hence $y = 0$ is an asymptote, and the curve is above it on each side.

When $x = 0$, $y = \frac{-4}{-12} = \frac{1}{3}$, therefore $\left(0, \frac{1}{3}\right)$ is on the curve.

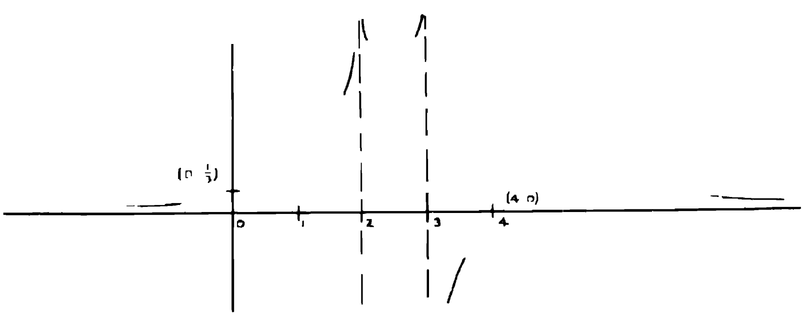


FIG. 27 (*a*).—Graph of $y = \frac{x-4}{(x-2)^2(x-3)}$

When $x = 2$ or 3, $y = \infty$, therefore $x = 2$ and $x = 3$ are asymptotes.

If x is just greater or just less than 2, y is always positive. The curve therefore goes to infinity, and returns, along the positive half of the asymptote $x = 2$.

If x is just less than 3, y is positive, while if x is just greater than 3, y is negative; the curve therefore goes up to infinity along the left-hand side of the asymptote $x = 3$, and goes down to "minus infinity" along the right-hand side.

The above fixes the parts of the curve shown in Fig. 27*a*; from these the rest of the curve can easily be joined up to give Fig. 27*b*.

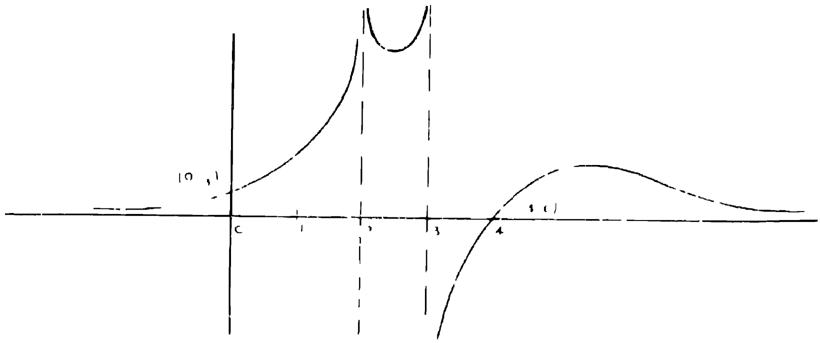


Fig. 27 (*b*) Graph of $y = \dfrac{x-4}{(x-2)^2(x-3)}$

Summary of equations of commonly occurring curves (*See* Fig. 28)

1. Straight line, parallel to x axis $\qquad y = k$
2. Straight line, of slope m, passing through ($x = 0$, $y = b$) $\qquad y = mx + b$
3. Straight line, of slope m, passing through ($x = a$, $y = 0$) $\qquad y = m(x - a)$
4. Straight line, passing through ($x = a$, $y = 0$) and ($x = 0$, $y = b$) $\qquad \dfrac{x}{a} + \dfrac{y}{b} = 1$
5. Circle, radius r, centre at the origin $\qquad x^2 + y^2 = r^2$
6. Circle, radius r, centre (a, b) $\qquad (x - a)^2 + (y - b)^2 = r^2$
7. Ellipse, semi-axes c and d, centre at the origin $\qquad \dfrac{x^2}{c^2} + \dfrac{y^2}{d^2} = 1$

8. Ellipse, semi-axes c and d, centre (a, b) with axes parallel to the co-ordinate axes :— $\frac{(x-a)^2}{c^2} + \frac{(y-b)^2}{d^2} = 1$

9. Parabola, vertex at the origin, with axis along x-axis :— $y^2 = 4ax$

10. Rectangular hyperbola, with axes as asymptotes, centre at the origin :— $xy = k^2$

11. Hyperbola, with foci on x-axis and centre at the origin :— $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$

12. General form of equation of conic section :— $ax^2 + 2hxy + by^2 + 2gx + 2fy + c = 0$

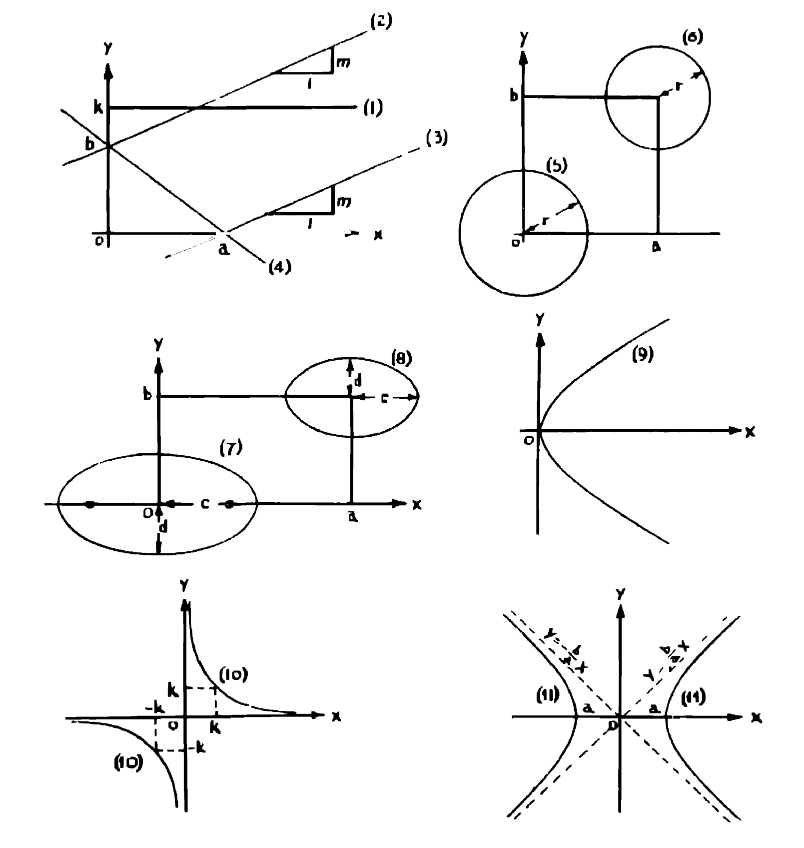


FIG. 28.—Commonly occurring curves.

TRIGONOMETRY

Angular measure

An angle is a measure of rotation; the usual units are degrees, where one degree is $\frac{1}{360}$th of a complete revolution. As a general rule, angles are measured anti-clockwise from a horizontal reference line, as in Fig. 29, where OA is the reference line, and θ is the angle AOP, shown thus $\angle AOP$.

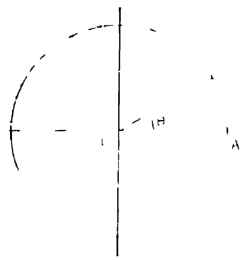


FIG. 29.—Angular measure.

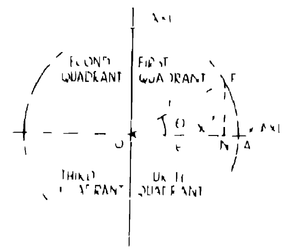


FIG. 30.—The four quadrants and the trigonometrical ratios.

A complete revolution is divided into 4 "quadrants" each of 90°, by two lines at right angles to one another passing through the centre.

In any circle, the length of the arc AP is directly proportional to $\angle AOP$. As the complete circumference $= 2\pi r$, then, by proportion,

$$\text{arc } AP = \frac{2\pi r\theta}{360} \tag{20}$$

where θ is in degrees.

The choice of degrees as units is inconvenient in many theoretical problems; the unit used in such cases is the "radian". One radian is defined as the angle subtended at the centre of a circle by an arc whose length is equal to the radius of the circle. Note that $360° = 2\pi$ radians; this gives 1 radian $= \frac{180}{\pi}$ degrees $=$ 57° 17′ 44″.

Using radians, the arc AP corresponding to an angle θ is :—

$$\text{arc } AP = r \times \theta \tag{21}$$

which is simpler than equation 20 using degrees for units.

The trigonometrical ratios

All right-angled triangles having one angle equal to θ will be similar, and the ratios of their sides will be equal. These ratios are very useful, and form the basis of trigonometry. The definitions of the various ratios hold for all angles, and are as follows.

Draw $\angle AOP$, and draw PN perpendicular to OA as in Fig. 30.

Let $OP = r$, $PN = y$, and $ON = x$.

The *sine* of the angle θ is then defined as the ratio of the length of PN to that of OP, *i.e.*, $\frac{y}{r}$; this is written as :—

$$\sin \theta = \frac{PN}{OP} = \frac{y}{r} \tag{22}$$

When P lies above the x-axis, y is positive, and therefore $\sin \theta$ is also positive; when P lies below the x-axis, y is negative, and therefore $\sin \theta$ is negative. Note that r is always considered positive.

The *cosine* of the angle θ is defined as the ratio of the length of ON to that of OP, *i.e.*, $\frac{x}{r}$, and is written as :—

$$\cos \theta = \frac{ON}{OP} = \frac{x}{r} \tag{23}$$

Note that $\frac{x}{r}$ is also equal to $\sin\left(\frac{\pi}{2} - \theta\right)$ so that :—

$$\cos \theta = \sin\left(\frac{\pi}{2} - \theta\right) \tag{24}$$

When P lies to the left of Oy, x is negative, and therefore $\cos \theta$ is also negative.

The *tangent* of the angle θ is defined as the ratio of the length of PN to that of ON, *i.e.*, $\frac{y}{x}$; then :—

$$\tan \theta = \frac{PN}{ON} = \frac{y}{x} \tag{25}$$

Note that :

$$\tan \theta = \frac{\sin \theta}{\cos \theta} \tag{26}$$

since

$$\frac{\sin \theta}{\cos \theta} = \frac{y/r}{x/r} = \frac{y}{x} = \tan \theta$$

Signs of the trigonometrical ratios.—If P is below Ox, PN (and therefore $\sin \theta$) will be negative, and if P is to the left of the vertical axis Oy, ON (and therefore $\cos \theta$) will be negative. Thus in the first quadrant, all the ratios are positive. In the second, only the sine is positive; in the third, only the tangent, and in the fourth, only the cosine (*see* Fig. 31). It is important to remember this when calculating ratios of angles outside the first quadrant.

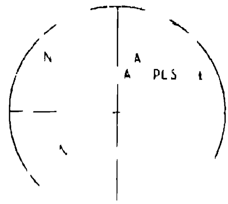


FIG. 31.—Signs of the ratios in the four quadrants.

Three other ratios are:—

$$\text{cosecant } \theta = \text{cosec } \theta = \frac{1}{\sin \theta} = \frac{r}{y} = \sec\left(\frac{\pi}{2} - \theta\right) \qquad (27)$$

$$\text{secant } \theta = \sec \theta = \frac{1}{\cos \theta} = \frac{r}{x} = \text{cosec}\left(\frac{\pi}{2} - \theta\right) \qquad (28)$$

$$\text{cotangent } \theta = \cot \theta = \frac{1}{\tan \theta} = \frac{x}{y} = \tan\left(\frac{\pi}{2} - \theta\right) \qquad (29)$$

Angles greater than 90° $\left(\frac{\pi}{2}\text{ radians}\right)$

Tables give the values of ratios for angles between 0° and 90°. To calculate ratios of an angle outside this range it must be written in the form 90° ± θ, 180° ± θ, 270° ± θ or 360° ± θ, where θ is an angle between 0° and 90°. One of the following relationships may then be used.

(*a*) The ratios of 180° ± θ, 360° ± θ, etc., are the same as the ratios of θ with a possible change of sign.

(*b*) For 90° ± θ, 270° ± θ, etc., sin becomes cos, cos becomes sin, tan becomes cot, and so on, with a possible change of sign.

(*c*) The ambiguity in (*a*) and (*b*) as regards the sign can be cleared up by considering the quadrant in which the original angle lies.

Example.—

Find $\sin (180° + \theta)$.

$\sin (180° + \theta)$ from (*a*) $= \pm \sin \theta$.

But $180° + \theta$ is in 3rd quadrant,

$\therefore \sin (180° + \theta)$ is negative,

$\therefore \sin (180° + \theta) = -\sin \theta$.

In practice this process is carried out mentally, thus:—

$$\cos (152°) = \cos (180° - 28°) = -\cos 28° = -0{\cdot}8829$$

or alternatively—

$$\cos (152°) = \cos (90° + 62°) = -\sin 62° = -0{\cdot}8829$$

This is illustrated in Fig. 32, where $OP = 1$.

Obviously, it is easier, if possible, to use the $180° \pm$ and $360° \pm$ forms than the $90° \pm$ and $270° \pm$ as they do not involve any change of ratio.

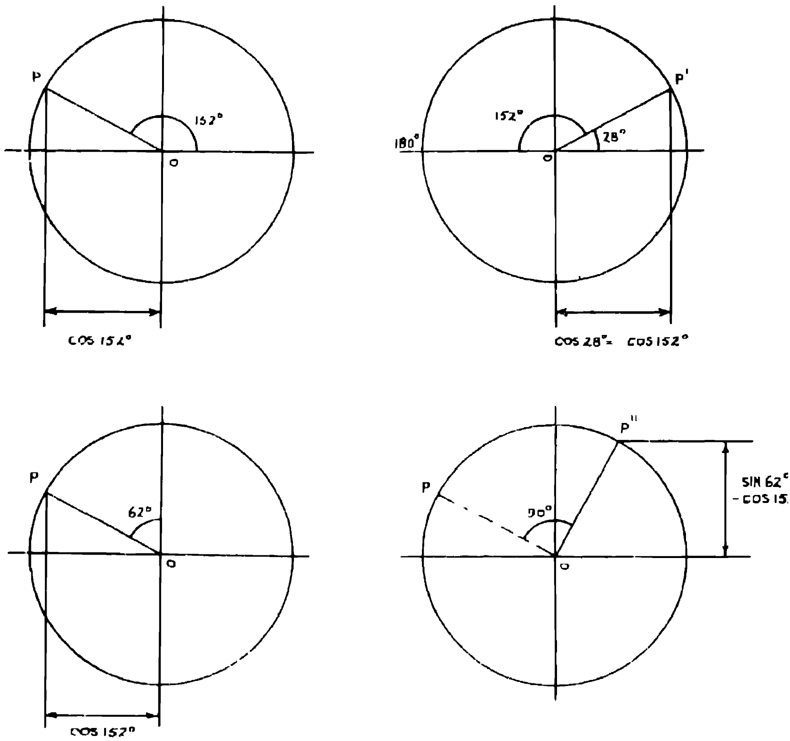


FIG. 32 - Examples of ratios outside the first quadrant.

The inverse functions

"The angle whose sine is x" is written as "$\sin^{-1} x$;" similarly, "$\cos^{-1} x$" means "the angle whose cosine is x". It is important not to confuse "$\sin^{-1} x$ the angle whose sine is x" with "$(\sin x)^{-1}$ $\left(\frac{1}{\sin x}\right)$". To avoid any possibility of confusion, the inverse ratios are sometimes written as "arc sin", "arc cos", *etc.*

Negative angles

Since θ is taken, by convention, as being measured in an anti-clockwise direction from the x-axis, "$-\theta$" represents an angle equal to θ, but measured clockwise from the x-axis, *i.e.* the same

is $(360° - \theta)$. Sin $(-\theta)$ is therefore equal to $-\sin\theta$,

$$\cos(-\theta) = +\cos\theta, \text{ and } \tan(-\theta) = -\tan\theta.$$

Particular angles

The ratios of angles such as 0°, 90°, 180°, *etc.*, also 30°, 45°, 60°, can be calculated from first principles.

For 0° and 90°, the limiting case of a triangle must be taken, when one side becomes zero and the other two both equal to the radius of the circle. Hence $\cos 0° = \sin 90° = 1$;

$$\sin 0° = \cos 90° = 0;\ \tan 0° = 0;\ \tan 90° = \infty.$$

45°.—If one angle of a right-angled triangle is 45°, the other must also be 45°; hence the triangle is isosceles—*i.e.*, the two sides adjacent to the right angle are equal. Hence by Pythagoras' theorem, each must be $\frac{1}{\sqrt{2}}$ of the hypotenuse (*see* Fig 33).

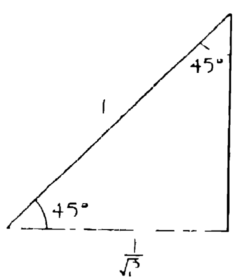


FIG. 33.—The trigonometrical ratios of 45°.

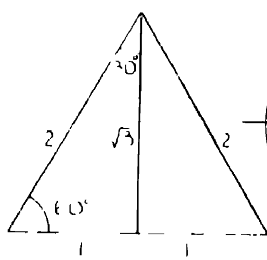


FIG. 34. The trigonometrical ratios of 30° and 60°.

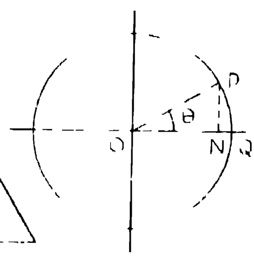


FIG. 35 — The trigonometrical ratios of very small angles

Thus $\quad \sin 45° = \cos 45° = \frac{1}{\sqrt{2}} = 0{\cdot}7071$, and $\tan 45° = 1$.

Similarly, by bisecting one of the angles of an equilateral triangle (*see* Fig. 34), it can be shown that:

$$\sin 30° = \cos 60° = \tfrac{1}{2} = 0{\cdot}5;\ \cos 30° = \sin 60° = \frac{\sqrt{3}}{2} = 0{\cdot}8660$$

$$\tan 30° = \frac{1}{\sqrt{3}} = 0{\cdot}5774 \qquad \tan 60° = \sqrt{3} = 1{\cdot}7321$$

Very small angles

Referring to Fig. 35, it can be seen that if θ be very small, the arc PQ ($= OP \,.\, \theta$) is roughly equal to PN ($= OP \,.\, \sin\theta$). Also ON is roughly equal to OP. For small angles, one can therefore take $\sin\theta$ and $\tan\theta$ as approximately equal to θ (where θ is expressed in radians), and $\cos\theta$ as approximately equal to 1. The smaller the angle, the more accurate is this approximation; the error is less than 2 per cent. for angles up to 20° for sines, and up to 13° for tangents.

General forms

The angles satisfying the equation —

$$\sin\theta = x$$

are given by $$\theta = n\pi + (-1)^n \sin^{-1} x$$

where n is any integer.

Similarly, the angles satisfying the equation

$$\cos\theta = x$$

are given by $$\theta = 2n\pi \pm \cos^{-1} x$$

And the angles satisfying the equation

$$\tan\theta = x$$

are given by $$\theta = n\pi + \tan^{-1} x$$

TABLE 1

The trigonometrical ratios for particular angles

Angle in radians	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$	$\frac{2\pi}{3}$	$\frac{3\pi}{4}$	$\frac{5\pi}{6}$	π	$\frac{3\pi}{2}$	2π
Angle in degrees	0	30	45	60	90	120	135	150	180	270	360
Sine	0	$\frac{1}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{\sqrt{3}}{2}$	$+1$	$\frac{\sqrt{3}}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{2}$	0	-1	0
Cosine	1	$\frac{\sqrt{3}}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{2}$	0	$-\frac{1}{2}$	$-\frac{1}{\sqrt{2}}$	$\frac{\sqrt{3}}{2}$	-1	0	$+1$
Tangent	0	$\frac{1}{\sqrt{3}}$	1	$\sqrt{3}$	∞	$-\sqrt{3}$	-1	$\frac{1}{\sqrt{3}}$	0	∞	0

Graphs of ratios

It is instructive at this stage to consider the graphs of the ratios; these are plotted in Fig. 36. Note that $\sin\theta$ and $\cos\theta$ lie between plus and minus one for all angles but that $\tan\theta$ assumes all values between plus and minus infinity.

Identities

Certain identities connecting the various trigonometrical ratios are important. They are

$$\sin^2\theta + \cos^2\theta = 1 \qquad (30)$$

$$\sec^2\theta = \tan^2\theta + 1 \qquad (31)$$

$$\operatorname{cosec}^2\theta = \cot^2\theta + 1 \qquad (32)$$

where θ may be any angle.

These may be proved from Pythagoras' theorem, for by definition —

$$\sin\theta = \frac{y}{r} \text{ and } \cos\theta = \frac{x}{r} \text{ (see Fig. 30)}$$

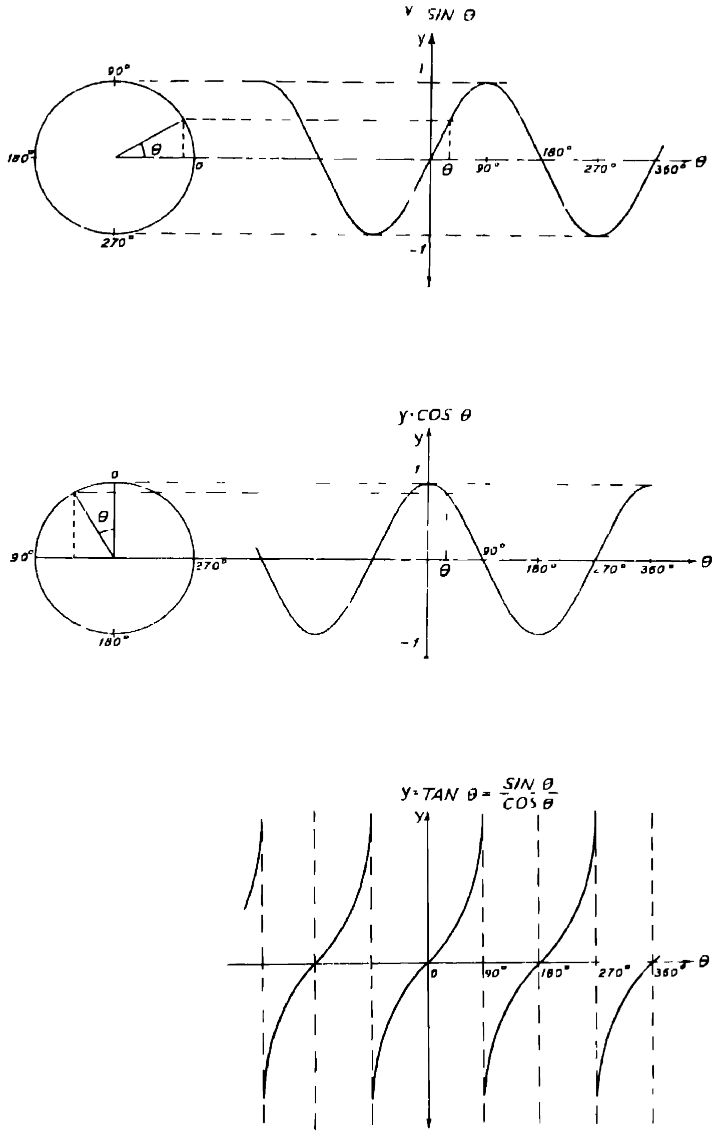


Fig. 36.—Graphs of trigonometrical ratios.

But x, y, r are the sides of a right-angled triangle,

$$\therefore \quad x^2 + y^2 = r^2$$

$$i.e., \quad r^2 \cos^2 \theta + r^2 \sin^2 \theta = r^2$$

$$i.e., \quad \cos^2 \theta + \sin^2 \theta = 1$$

The other identities are obtained by dividing both sides of this equation by $\cos^2 \theta$ and $\sin^2 \theta$ respectively.

Multiple angle formulae

It is often useful to be able to express the trigonometrical ratios of the sum of two angles in terms of the ratios of each individual angle. This may be done as follows: it will be proved that —

$$\sin(A + B) = \sin A \cos B + \cos A \sin B$$

(See Fig. 37)

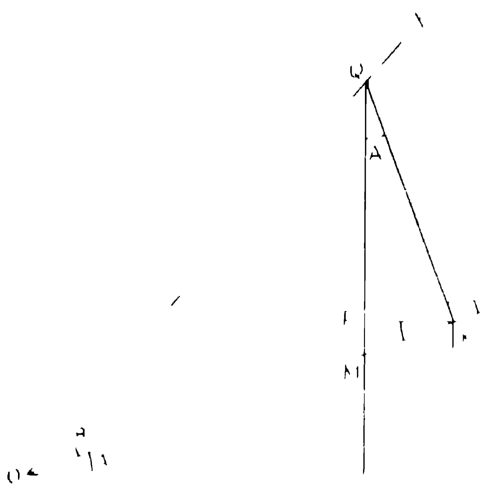


Fig. 37. Illustrating formula for $\sin(A + B)$.

Draw — $\angle YOZ = A$

$\angle XOY = B$

So $\angle XOZ = A + B$

Q is any point on OX

Draw QK perpendicular to OZ cutting OY in M

„ QR „ „ OY

„ RN „ „ OZ

„ RP „ „ OK

As $\angle K = 90°$, $\angle OMK = 90° - A = \angle QMR$

As $\angle QRM = 90°$, $\angle MQR = 90° - \angle QMR = 90° - (90° - A) = A$

Similarly, it will be seen that $\angle PRM = A$

$$\sin(A+B) = \frac{QK}{OQ} = \frac{QP}{OQ} + \frac{PK}{OQ}$$

But $\quad QP = QR\cos \angle PQR = QR\cos A$

and $\quad PK = RN = OR\sin \angle NOR = OR\sin A$

$$\therefore \quad \sin(A+B) = \frac{QR}{OQ}\cos A + \frac{OR}{OQ}\sin A$$

But $\quad \dfrac{QR}{OQ} = \sin B \quad$ and $\quad \dfrac{OR}{OQ} = \cos B$

$$\therefore \quad \sin(A+B) = \sin A\cos B + \cos A\sin B \tag{33}$$

Similarly it may be shown that

$$\sin(A-B) = \sin A\cos B - \cos A\sin B \tag{34}$$
$$\cos(A+B) = \cos A\cos B - \sin A\sin B \tag{35}$$
$$\cos(A-B) = \cos A\cos B + \sin A\sin B \tag{36}$$

By adding (33) and (34)

$$2\sin A\cos B = \sin(A+B) + \sin(A-B) \tag{37}$$

By subtracting (34) from (33)

$$2\cos A\sin B = \sin(A+B) - \sin(A-B) \tag{38}$$

By adding (35) and (36)

$$2\cos A\cos B = \cos(A+B) + \cos(A-B) \tag{39}$$

By subtracting (35) from (36)

$$2\sin A\sin B = \cos(A-B) - \cos(A+B) \tag{40}$$

(37), (38), (39) and (40) are useful for expressing the product of two ratios as the sum of two ratios.

By letting $A = \dfrac{C+D}{2}$ and $B = \dfrac{C-D}{2}$, these equations become :—

$$\sin C + \sin D = 2\sin\frac{C+D}{2}\cos\frac{C-D}{2} \tag{41}$$

$$\sin C - \sin D = 2\cos\frac{C+D}{2}\sin\frac{C-D}{2} \tag{42}$$

$$\cos C + \cos D = 2\cos\frac{C+D}{2}\cos\frac{C-D}{2} \tag{43}$$

$$\cos C - \cos D = 2\sin\frac{C+D}{2}\sin\frac{D-C}{2} = -2\sin\frac{C+D}{2}\sin\frac{C-D}{2} \tag{44}$$

These formulae are useful for expressing the sum of two ratios as the product of two ratios.

The formulae for $\tan (A \pm B)$ are :—

$$\tan (A + B) = \frac{\tan A + \tan B}{1 - \tan A \tan B} \tag{45}$$

$$\tan (A - B) = \frac{\tan A - \tan B}{1 + \tan A \tan B} \tag{46}$$

Double angle formulae.

By putting $A = B$ in the above, the following expressions are derived for ratios of $2A$:—

$$\sin 2A = 2 . \sin A . \cos A \tag{47}$$

$$\cos 2A = \cos^2 A - \sin^2 A \tag{48}$$

$$\left.\begin{aligned} &= 2 \cos^2 A - 1 \\ &= 1 - 2 \sin^2 A \end{aligned}\right\} \text{ since } \cos^2 A + \sin^2 A = 1$$

$$\tan 2A = \frac{2 \tan A}{1 - \tan^2 A} \tag{49}$$

The following are useful forms of (48) :—

$$\sin^2 A = \tfrac{1}{2} (1 - \cos 2A) \tag{50}$$

$$\cos^2 A = \tfrac{1}{2} (1 + \cos 2A) \tag{51}$$

Relationships of this nature facilitate the solution of many problems encountered in telecommunication engineering; this is illustrated by the following examples, which prove three of the relationships used later in this book.

Example 1.—

Express $A \sin \omega t + B \cos \omega t$ in the form :

$$r . \sin (\omega t + \theta)$$

Let :— $A = r . \cos \theta$

and :— $B = r . \sin \theta$

Squaring and adding these two equations,

$$A^2 + B^2 = r^2$$

$$\therefore \quad r = \sqrt{A^2 + B^2}$$

Dividing the second equation by the first,

$$\frac{B}{A} = \tan \theta$$

Then

$$\begin{aligned} A \sin \omega t + B \cos \omega t &= r . \cos \theta . \sin \omega t + r . \sin \theta . \cos \omega t \\ &= r . \sin (\omega t + \theta) \\ &= \sqrt{A^2 + B^2} . \sin (\omega t + \theta) \quad \textit{Ans.} \end{aligned}$$

$$\text{where } \theta = \tan^{-1} \frac{B}{A}$$

Example 2.—

Prove that $\tan^{-1} A + \tan^{-1} B = \tan^{-1} \left(\dfrac{A + B}{1 - AB} \right)$

Put $x = \tan^{-1} A$, that is, $\tan x = A$

and $y = \tan^{-1} B$, that is, $\tan y = B$

Then

$$\tan(\tan^{-1} A + \tan^{-1} B) = \tan(x + y)$$

$$= \frac{\tan x + \tan y}{1 - \tan x \tan y}$$

$$= \frac{A + B}{1 - AB}$$

$$\therefore \quad \tan^{-1} A + \tan^{-1} B = \tan^{-1}\left(\frac{A + B}{1 - AB}\right) \qquad \text{Q.E.D.}$$

Similarly it can be proved that :—

$$\tan^{-1} A - \tan^{-1} B = \tan^{-1}\left(\frac{A - B}{1 + AB}\right)$$

Example 3.—Prove that :—

$$\tan^{-1}\frac{1}{A} - \tan^{-1}\frac{1}{B} = \tan^{-1} B - \tan^{-1} A$$

Let :— $A = \tan\theta$, so that $\theta = \tan^{-1} A$

Then :— $\frac{1}{A} = \frac{1}{\tan\theta} = \cot\theta$

$$= \tan\left(\frac{\pi}{2} - \theta\right)$$

$$\therefore \tan^{-1}\frac{1}{A} = \frac{\pi}{2} - \theta$$

$$= \frac{\pi}{2} - \tan^{-1} A$$

Similarly $\tan^{-1}\frac{1}{B} = \frac{\pi}{2} - \tan^{-1} B$

Hence

$$\tan^{-1}\frac{1}{A} - \tan^{-1}\frac{1}{B} = \tan^{-1} B - \tan^{-1} A$$

Triangle formulae

These apply to any triangle having angles A, B, C, and sides a, b, c (*see* Fig. 38).

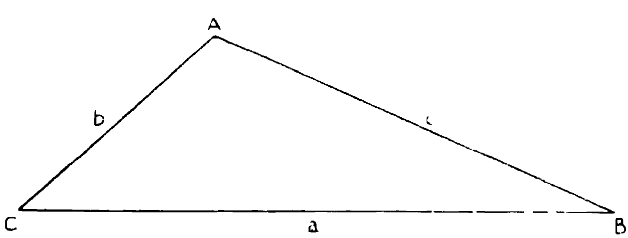


FIG. 38.—Triangle with sides *abc* and angles *ABC*.

$$\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C} \quad \text{(The "sine rule")} \tag{52}$$

$$a^2 = b^2 + c^2 - 2bc \cos A \quad \text{(The "cosine rule")} \tag{53}$$

$$b^2 = c^2 + a^2 - 2ca \cos B$$

$$c^2 = a^2 + b^2 - 2ab \cos C$$

$$\tan\left(\frac{A-B}{2}\right) = \frac{a-b}{a+b} \cot \frac{C}{2} \quad \textit{etc.} \tag{54}$$

$$\sin \frac{A}{2} = \sqrt{\frac{(s-b)(s-c)}{bc}} \quad \textit{etc.} \tag{55}$$

$$\cos \frac{A}{2} = \sqrt{\frac{s(s-a)}{bc}} \quad \textit{etc.} \tag{56}$$

$$\sin A = \frac{2}{bc}\sqrt{s(s-a)(s-b)(s-c)} \quad \textit{etc.} \tag{57}$$

where $s = \frac{1}{2}(a + b + c)$

Example.

Given that, in the triangle of Fig. 38, two sides and the included angle are known, find the remaining side and angles, *e.g.*, given $a = 20$, $b = 10$, $C = 40°$, find c, A and B.

Applying the cosine rule (equation 53):

$$\begin{aligned} c^2 &= a^2 + b^2 - 2ab \cos C \\ &= 20^2 + 10^2 - 2 \cdot 20 \cdot 10 \cos 40° \\ &= 400 + 100 - 400 \times 0{\cdot}7660 \\ &= 500 - 306{\cdot}4 \\ &= 193{\cdot}6 \end{aligned}$$

i.e., $c = 13{\cdot}91$. *Ans.* (*i*)

Now applying the sine rule (equation 52):

$$\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C}$$

i.e.,
$$\frac{20}{\sin A} = \frac{10}{\sin B} = \frac{13{\cdot}91}{\sin 40°}$$

$$\therefore \quad \sin A = \frac{20 \sin 40°}{13{\cdot}91} = \frac{20 \times 0{\cdot}6428}{13{\cdot}91} = 0{\cdot}9241$$

$$\therefore \quad A = 67°\ 32' \text{ or } 112°\ 28'$$

Also
$$\sin B = \frac{10 \sin 40°}{13{\cdot}91} = 0{\cdot}4622$$

$$\therefore \quad B = 27°\ 32' \text{ or } 152°\ 28'$$

Thus it would appear that there are two triangles having the given properties; but this is not so for the angles A, B and C of the

triangle must satisfy the relationship —

$$A + B + C = 180°$$

and the only combination of the above angles which satisfies this result is —

$$A = 112° 28', \; B = 27° 32' \quad Ans \; (ii \;\&\; iii)$$

A simpler approach for finding B having determined two possible values of A is as follows

$$B = 180 - (A + C)$$

$$\therefore \quad B = 180 - (67° 32' + 40°)$$

$$\text{or} \quad = 180 - (112° 28' + 40°)$$

In this case the correct pair of values for A and B may be selected by means of the self-evident result that if a triangle has two unequal angles the greater angle will be opposite the greater side.

Since $b < c$ it follows that $B < C$

$$B = 27° 32' \text{ and } A = 112° 28' \quad Ans \; (iii \;\&\; ii)$$

The triangle is therefore uniquely determined as in Fig 39

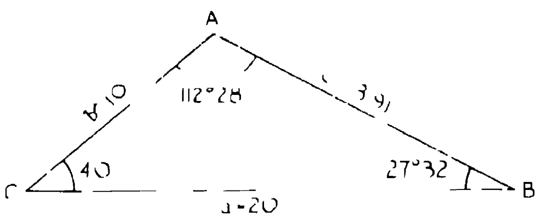


Fig 39 Triangle ... Example

An alternative approach to the same problem is to employ equation 51

$$\tan\left(\frac{A - B}{2}\right) = \frac{a - b}{a + b}\cot\frac{C}{2}$$

$$= \frac{20 - 10}{20 + 10}\cot\frac{40}{2}$$

$$= \tfrac{1}{3}\cot 20 = 0.9158$$

$$\therefore \quad \frac{A - B}{2} = 42° 29' \text{ (no ambiguity)}$$

Put $$\frac{A + B}{2} = \frac{180 - C}{2} = 70$$

Hence adding and subtracting gives

$$A = 112° 29' \text{ and } B = 27° 31' \quad Ans \; (ii \;\&\; iii)$$

Applying the sine rule

$$c = \frac{a \sin C}{\sin A}$$

$$\therefore \quad c = 13.91 \quad Ans \; (i)$$

BINOMIAL THEOREM

Expansions such as
$$(1 + x)^2 = 1 + 2x + x^2$$
$$(1 + x)^3 = 1 + 3x + 3x^2 + x^3$$
are often useful. A general form of this expansion is given by the Binomial theorem :—

$$(1 + x)^n = 1 + nx + \frac{n(n-1)}{1.2}x^2 + \frac{n(n-1)(n-2)}{1.2.3}x^3 + \ldots \qquad (58)$$

If n is a positive integer, this is true for all values of x. It can be seen that in this case the expansion is finite, the last term being x^n. (The expansions for $(1+x)^2$ and $(1+x)^3$ may be verified.)

If n is not a positive integer, the expansion contains an infinite number of terms. It is then valid only if $-1 < x < 1$ (*i.e.*, $|x| < 1$). In this case a series having an infinite number of terms has a finite sum. Such a series is said to be "convergent".

Factorials. —The terms of this expansion —as well as of many others—can be simplified by using the "factorial" notation. "Factorial n" is written as $\underline{|n}$ or $n!$, and is defined as :—

$$\underline{|n} = n \times (n-1) \times (n-2) \times \ldots \times 4 \times 3 \times 2 \times 1 \qquad (59)$$

Thus
$$\underline{|6} = 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$$
$$\underline{|3} = 3 \times 2 \times 1 = 6$$
$$\underline{|1} = 1$$

Using this notation, the Binomial expansion of equation 58 can be rewritten as :—

$$(1 + x)^n = 1 + nx + \frac{n(n-1)}{\underline{|2}}x^2 + \frac{n(n-1)(n-2)}{\underline{|3}}x^3 + \ldots \qquad (60)$$

or alternatively :—

$$(1 + x)^n = 1 + \frac{\underline{|n}}{\underline{|n-1}}x + \frac{\underline{|n}}{\underline{|2}\,\underline{|n-2}}x^2 + \frac{\underline{|n}}{\underline{|3}\,\underline{|n-3}}x^3 + \ldots \qquad (61)$$

Example. - Expand $(1 + x)^5$

$$(1 + x)^5 = 1 + 5x + \frac{5.4}{1.2}x^2 + \frac{5.4.3}{1.2.3}x^3 + \frac{5.4.3.2}{1.2.3.4}x^4 + \frac{5.4.3.2.1}{1.2.3.4.5}x^5$$
$$= 1 + 5x + 10x^2 + 10x^3 + 5x^4 + x^5$$

This may be verified by putting $x = 1$

$$\therefore\ 2^5 = 1 + 5 + 10 + 10 + 5 + 1$$
$$= 32, \text{ which is true.}$$

Exponential series

This important series can be derived from the Binomial theorem.

Consider the expansion of $\left(1 + \frac{1}{n}\right)^{nx}$.

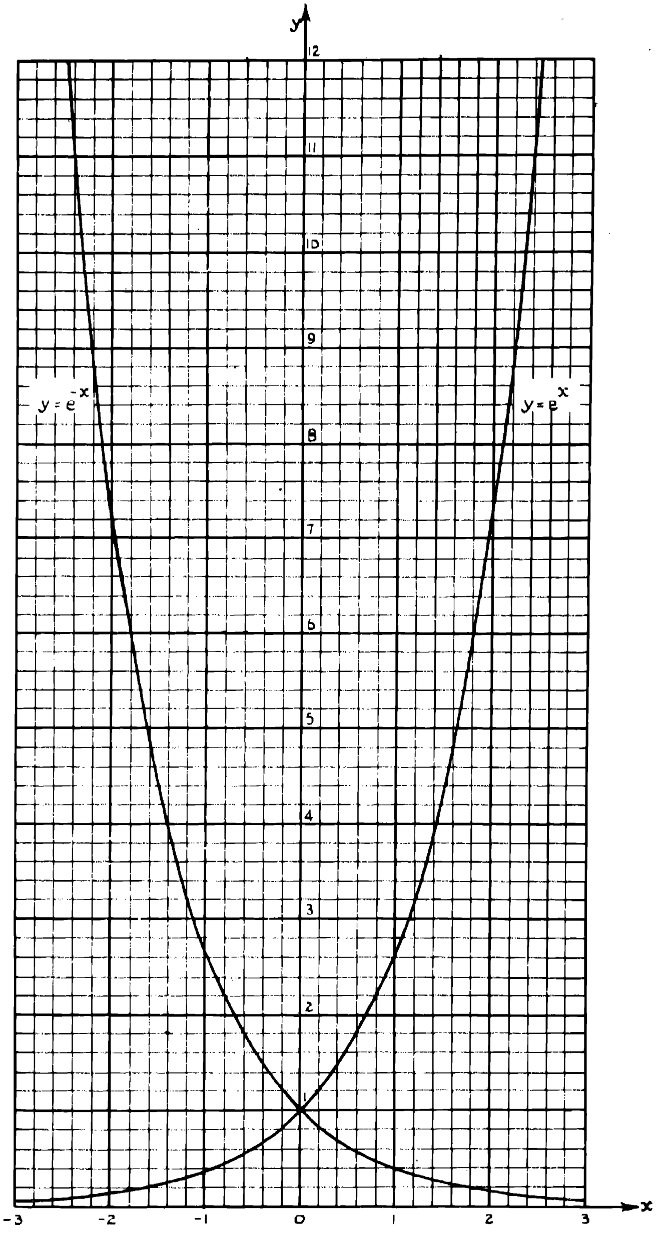


FIG. 40.—Graphs of e^x and e^{-x}.

The r^{th} term is $\dfrac{(nx)(nx-1)(nx-2)\ldots(nx-r+1)}{\lfloor r} \times \dfrac{1}{n^r}$

If n is made very large each bracket in the numerator $\simeq nx$

Therefore as $n \to \infty$ the r^{th} term $\to \dfrac{x^r}{\lfloor r}$

i.e. when $n \to \infty$ $\left(1+\dfrac{1}{n}\right)^{nx} = 1 + x + \dfrac{x^2}{\lfloor 2} + \ldots$ to infinity (62)

or $\underset{n \to \infty}{\text{Limit}} \left[\left(1+\dfrac{1}{n}\right)^n\right]^x = 1 + x + \dfrac{x^2}{\lfloor 2} + \ldots$

It can be shown that this infinite series is convergent for all values of x.

The last equation can be written as

$$e^x = 1 + x + \frac{x^2}{\lfloor 2} + \ldots \quad (63)$$

where

$$e = \underset{n \to \infty}{\text{Limit}} \left(1+\frac{1}{n}\right)^n \quad (64)$$

The value of e can be calculated by letting $x = 1$

$$e = e^1 = 1 + 1 + \frac{1}{\lfloor 2} + \frac{1}{\lfloor 3} + \ldots \quad (65)$$
$$= 2.71828$$

This expansion for e^x is very useful. When dealing with it it is important to remember that e^x is an ordinary number raised to the power x and behaves just as any other number. The usefulness of the expansion is that it enables a number raised to a power to be expressed as a series.

The graphs of $y = e^x$ and $y = e^{-x}$ are important; they are given in Fig. 40.

Logarithms to the base e are very important and are known as Natural or Napierian logarithms. When no base is stated, it may be assumed to be e. Thus $\log X$ stands for logarithm of X to the base e.

$$\log_e X = \frac{\log_{10} X}{\log_{10} e}$$

But $\log_{10} 2.718 = 0.4343$

Thus $\log_e X = \dfrac{\log_{10} X}{0.4343}$

i.e. $\log_e X = 2.3026 \log_{10} X$ (66)

and $\log_{10} X = 0.4343 \log_e X$ (67)

Logarithms both to the base 10 and to the base e may be found in Appendix I.

VECTORS

Definition

A vector is a quantity *possessing both magnitude and direction*. It may be represented by a line whose length and angle with some reference axis correspond to this magnitude and direction. A scalar, on the other hand, is *a quantity that possesses magnitude alone*. Thus if a man walks four miles East from O to P and then three miles North from P to Q, the total distance he has covered

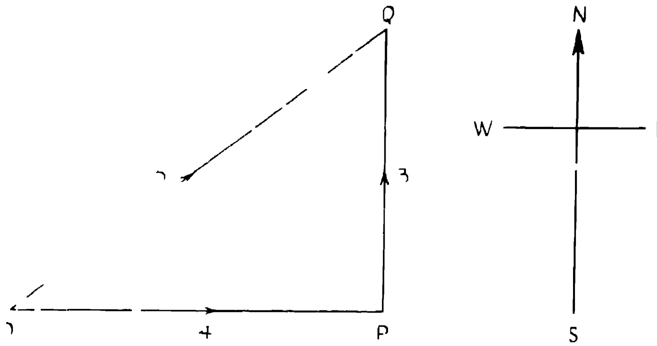


Fig. 41. Example of vector addition.

is $OP + PQ = 4 + 3 = 7$ miles, a *scalar* quantity. But the distance OQ from his starting point to his destination is 5 miles in the direction shown in Fig. 41. This is a *vector* quantity since it has *direction* as well as magnitude. When it is necessary to emphasize the distinction between scalar and vector quantities, an arrow may be placed over the two letters representing its origin and termination in order to indicate its direction. Thus, referring to Fig. 41, one could write

$$\vec{OP} + \vec{PQ} = \vec{OQ}$$

in place of vector OP + vector PQ = vector OQ.

Addition of vectors

The sum of two vectors is defined as the displacement equivalent to the combined effect of the two individual displacements. Thus

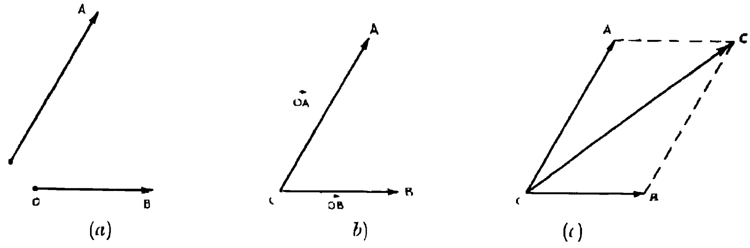


Fig. 42. Addition of two vectors OA and OB

in Fig. 41, the sum of the vectors OP and PQ is the vector OQ. In general, two vectors OA and $O'B'$ (*see* Fig. 42*a*) are added by drawing them so that they both emanate from one point, —*i.e.* as $\vec{OA}$ and $\vec{OB}$ in Fig. 42*b* The sum of these two, $\vec{OA} + \vec{OB}$, is the vector OC, which may be obtained either by drawing AC equal in direction and magnitude to OB, or by drawing BC equal in direction and magnitude to OA. In either case, the same point C is arrived at (Fig. 42*c*), and it can be seen that the figure $OBCA$ is a parallelogram.

This gives the rule for addition of vectors :—

The resultant of two vectors OA and OB is the diagonal through O of the parallelogram having OA and OB as two sides.

To find the magnitude of the resultant OC and the angle θ_1 that it makes with vector OB, drop a perpendicular CD on to OB, as in Fig. 43.

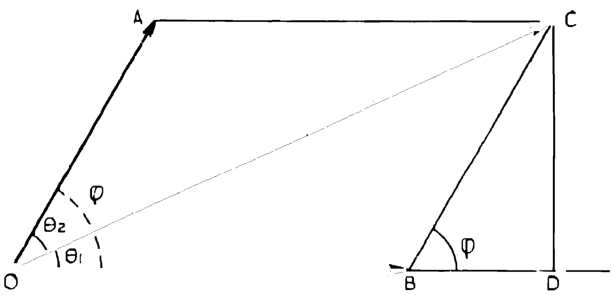


Fig. 43.—Parallelogram for addition of vectors.

Let the angle between vectors OA and OB be φ.

As OA and BC are parallel,

$$\angle DBC = \varphi$$

$$\angle OBC = 180° - \varphi$$

Applying the cosine rule to triangle OBC :—

$$OC^2 = BC^2 + OB^2 - 2BC \,.\, OB \cos \angle OBC$$

i.e. $$OC^2 = OA^2 + OB^2 + 2OA \,.\, OB \cos \varphi$$

$$\therefore \quad OC = \sqrt{OA^2 + OB^2 + 2\,OA \,.\, OB \cos \varphi} \tag{68}$$

$$\tan \theta_1 = \frac{CD}{OD}$$

$$= \frac{BC \sin \varphi}{OB + BD}$$

$$= \frac{OA \sin \varphi}{OB + BC \cos \varphi} = \frac{OA \sin \varphi}{OB + OA \cos \varphi} \tag{69}$$

Similarly if θ_2 is the angle between the resultant OC and the vector OA —

$$\tan\theta_2 = \frac{OB\sin\varphi}{OA + OB\cos\varphi} \tag{70}$$

Negative vectors

To comply with the normal use of + and − signs the vector "$-OA$" is defined as that vector which when added to OA will produce a zero resultant. (A zero vector has zero magnitude.)

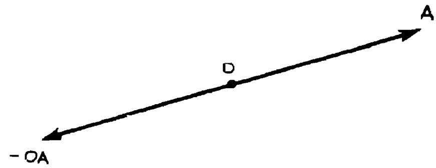


Fig. 44—Negative vector OA

From this definition it follows that the vector $-\vec{OA}$ has the same magnitude as $\vec{OA}$ and the opposite direction (see Fig. 44) so that $-\vec{OA} = \vec{AO}$.

This is equivalent to saying that the minus sign rotates the vector through 180°.

Subtraction of vectors

The result of subtracting a vector OB from a vector OA can be obtained by regarding $\vec{OA} - \vec{OB}$ as $\vec{OA} + (-\vec{OB})$ as in Fig. 45

$$\vec{OA} - \vec{OB} = \vec{OA} + (-\vec{OB}) = \vec{OC}$$

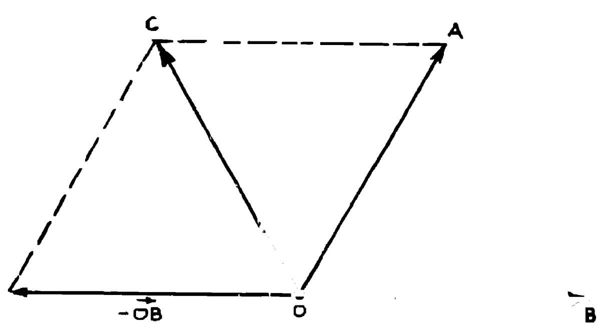


Fig. 45—Subtraction of vector OB from OA

Multiplication of a vector by a number

Multiplication has not yet been defined except that multiplication by "-1" rotates a vector through 180°. If a vector $\vec{OP}$ be multiplied by any positive real number n the result is a vector $\vec{OP'}$ of length equal to ($n \times OP$) and of the same direction as $\vec{OP}$. Thus if $\vec{OP}$ is 4 miles eastwards, $3 \times \vec{OP}$ is a distance of twelve miles eastwards (see Fig. 46b). If a vector be multiplied by a negative number its direction is reversed since $-n = -1 \times n$. Thus -3 times the vector $\vec{OP}$ previously mentioned is the vector $\vec{OP'}$ which is twelve miles westwards (Fig. 46c).

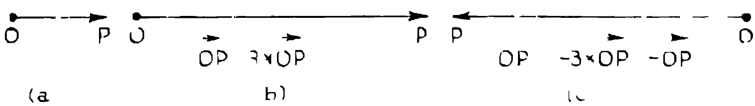


Fig. 46.—Multiplication of a vector.

Rotation of a vector

Since multiplication of a vector by -1 rotates it through 180° it can be seen that multiplying it *twice* by -1 will rotate it through 360°. In general multiplication by $(-1)^n$ rotates a vector through $(n \times 180°)$ if n is an integer. If this rule is to apply for *all* values of n then on putting $n = \frac{1}{2}$ multiplication by $(-1)^{\frac{1}{2}}$ must be considered to rotate a vector through $\frac{1}{2} \times 180° = 90°$. $(-1)^{\frac{1}{2}} = \sqrt{-1}$ is denoted by the letter j. The direction taken by convention as positive is anticlockwise. Multiplication of a vector by j thus rotates it anticlockwise through 90°.

$\sqrt{-1}$ cannot be evaluated in terms of normal numbers and it is known as an "imaginary" quantity. It can however be

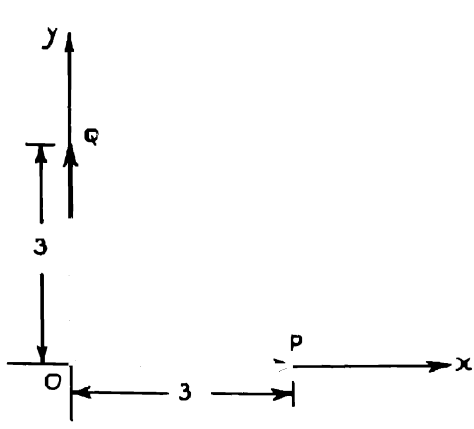


Fig. 47.—Rotation of a vector $OQ = jOP$

dealt with is a normal algebraic quantity. As will be seen, it frequently permits great simplification in calculations.

The important result of the last paragraph is the fact that multiplication by $j = \sqrt{-1}$ rotates a vector anticlockwise through 90°. It follows that multiplication of a vector by $j^2 = -1$ rotates it through 180°, by $j^3 = -j = \frac{1}{j}$ through 270°, and by $j^4 = 1$ through 360°.

This last result provides a convenient way of representing vectors when numerical problems are concerned.

Draw two rectangular axes Ox, Oy, and let the series of real numbers, positive or negative, represent vectors along the x axis, e.g. the number 3 represents a vector of length 3 along Ox (see Fig. 47, where OP represents the vector 3).

Consider now a purely imaginary number such as $j3$. This can be written as $j \times 3$. But the vector 3 is a vector along the x axis of length 3, and multiplication by j rotates the vector through 90°. Therefore the vector $j3$ is a vector of length 3 along OY, i.e. OQ represents the vector $j3$.

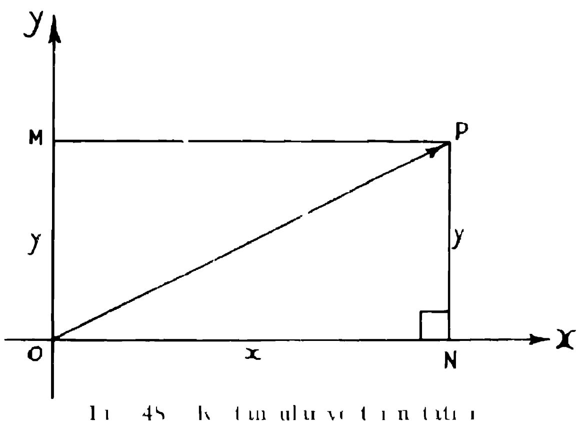


Fig. 48. Rectangular vector notation

Rectangular notation

Consider the vector OP (Fig. 48) and drop perpendiculars PN, PM from P to the axes. Let $ON = x$, $OM = y$. Then (x, y) are the coordinates of P. The vector OP is the sum of the vectors ON, OM (since $ONPM$ is a parallelogram). But ON is of length x and is along the x axis, therefore it is equal to the vector x. Also OM has length y and is along the y axis, and it is consequently equal to the vector jy. Therefore one can write:—

$$\text{Vector } OP = ON + OM = x + jy \qquad (71)$$

Hence *any* vector OP may be written in the form $(x + jy)$, where x and y are the coordinates of its extremity P. For example, in

Fig. 49, the vectors OA and OB are denoted as the vectors $1 + j3$ and $2 - j2$ respectively.

This method of notation simplifies the addition of vectors, for if the vectors are all in the form $x + jy$ (*i.e.*, real + imaginary terms), the resultant of adding will be found by adding all the real terms and adding all the imaginary terms. For example, in Fig. 49, the sum of $OA = 1 + j3$ and $OB = 2 - j2$ will be $OC = 3 + j1$.

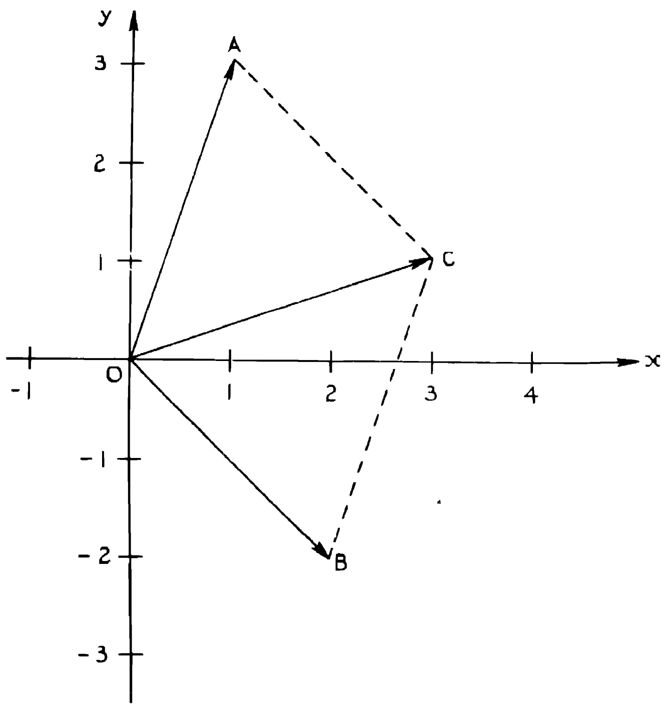


FIG. 49.—Addition of vectors by components.

This may be verified by drawing the parallelogram. Any number of vectors may be added or subtracted in this way.

For a vector to be zero, both the real (x) and the imaginary (y) terms must be equal to 0, and for two vectors to be equal, the real parts of the two vectors must be equal, as also must the imaginary.

Polar notation (modulus and angle)

It has been shown that any vector may be represented by its real and imaginary parts. Alternatively, a vector may be denoted by its length—known as its "modulus"—and the angle it makes with the x-axis—sometimes known as its "argument". Using this notation, if the vector OP has length r and makes an angle θ with

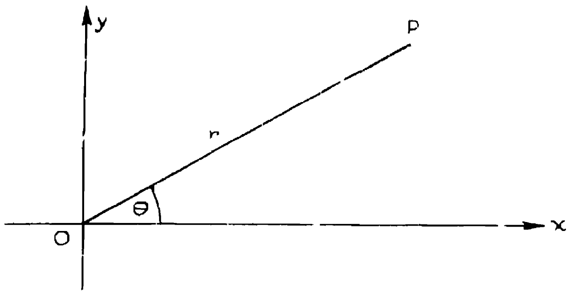


FIG. 50.—Polar vector notation.

the x axis it is called the vector $r \angle \theta$ (see Fig. 50).

$$\text{i.e.} \qquad \text{Vector } OP = r \angle \theta \qquad (72)$$

It should be noted that the modulus of OP is often written as $|OP|$, i.e. $|OP| = r$.

The negative of OP (that is, the vector $-OP$) has the same length (r) as OP but the reverse direction. It can therefore be expressed as

$$-OP = -r \angle \theta$$

$$= r \angle \pi + \theta \qquad (73)$$

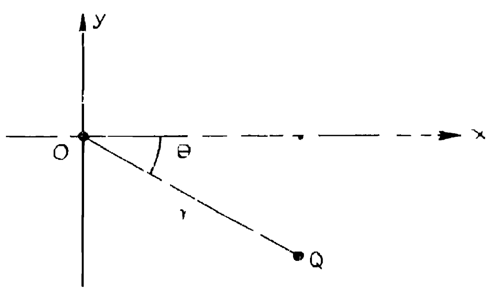


FIG. 51.—Vector with negative angle.

The vector OQ (see Fig. 51) is the vector $r \angle -\theta$. This is sometimes written as the vector $r \angle \underline{\theta}$ where $\angle -\theta = \angle \underline{\theta}$.

The vector $r \angle \theta$ should not be confused with the vector $-r \angle \underline{\theta}$.

Thus vector $3 \angle -45 = 3 \angle \underline{45} = 3 \angle 315$,

whereas vector $-3 \angle \underline{45} = 3 \angle \underline{180 + 45°} = 3 \angle \underline{225}$.

Conversion from rectangular notation to polar notation

Using the rectangular notation the vector OP (Fig. 52) is denoted as $x + jy$; while using the polar notation, it is denoted as $r \angle \theta$.

Clearly, simple relationships exist between x, y, r and θ. Using Pythagoras' theorem :—

$$|OP|^2 = r^2 = x^2 + y^2$$

$$\therefore \quad |OP| = r = \sqrt{x^2 + y^2} \tag{74}$$

e.g., if $OP = 3 + j4$, then $r = \sqrt{9 + 16} = \sqrt{25} = 5$

It is important to remember that $r = \sqrt{x^2 + y^2}$ *and not* $\sqrt{x^2 + (jy)^2}$.

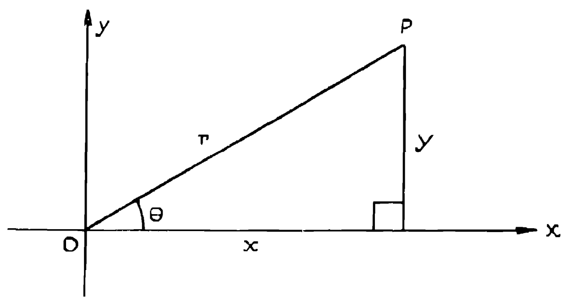


FIG. 52.—Relation between rectangular and polar notation.

The angle θ is best given by the formula :—

$$\tan \theta = \frac{y}{x} \tag{75}$$

This gives θ with a possible error of 180°; the ambiguity can be overcome by considering in which quadrant the vector lies.

e.g., if $\qquad OP = 3 + j4, \ \tan\theta = \frac{4}{3} = 1\cdot 33.$

This gives $\theta = 53° 7'$ or $180° + 53° 7'$; as the vector is in the 1st quadrant, the first result is correct, *i.e.*, $\theta = 53° 7'$.

Example.—

Express in polar form the vector :—

$$OP = a + j\sqrt{1 - a^2}$$

$$|OP| = \sqrt{a^2 + (1 - a^2)} = 1$$

$$\tan \varphi = \frac{\sqrt{1 - a^2}}{a}$$

Hence the vector $a + j\sqrt{1 - a^2}$ may also be denoted as the vector :

$$1 \angle \tan^{-1} \frac{\sqrt{1 - a^2}}{a}. \quad \textit{Ans.}$$

Conversion from polar notation to rectangular notation

If the vector is given in the form $r \underline{/\theta}$, it may easily be rewritten in the rectangular form as $x + jy$.

Referring to Fig. 52 :—

$$x = r \cos \theta$$
$$y = r \sin \theta \qquad (76)$$
$$\therefore \qquad OP \equiv r \underline{/\theta} = x + jy \qquad (77)$$
$$= r \cos \theta + j\, r \sin \theta$$
$$= r (\cos \theta + j \sin \theta) \qquad (78)$$

It is important to note that the form of the vector $r \underline{/\theta}$ *is* "$r (\cos \theta + j \sin \theta)$", *and not* "$r (\sin \theta + j \cos \theta)$". *The latter may be rewritten as* $r [\cos (90 - \theta) + j \sin (90 - \theta)]$, *and it is therefore the vector* $r \underline{/90^\circ - \theta}$.

Example.—
Write the vector $OP = 2 \underline{/30^\circ}$ in the form $x + jy$.

$$OP = r \underline{/\theta} = 2 \underline{/30^\circ} = 2 \cos 30^\circ + j \,.\, 2 \sin 30^\circ$$
$$= \sqrt{3} + j\, 1. \quad \textit{Ans.}$$

Multiplication and division of vectors

The process of multiplication using rectangular notation is purely algebraic.

e.g. $\qquad (3 + j4)(1 - j1) = 3 - j3 + j4 - j^2 4$
$$= 3 - j3 + j4 + 4$$
$$= 7 + j\, 1.$$

In this form, it is not so easy to see what the result means as when the vectors are expressed in polar form; nor, as will be seen, is multiplication of vectors as easy when they are expressed in the rectangular form as when they are expressed in polar form.

Consider the two vectors $OP_1 = r_1 \underline{/\theta_1}$ and $OP_2 = r_2 \underline{/\theta_2}$.

Then $\qquad OP_1 \equiv r_1 \underline{/\theta_1} = r_1 (\cos \theta_1 + j \sin \theta_1)$

and $\qquad OP_2 \equiv r_2 \underline{/\theta_2} = r_2 (\cos \theta_2 + j \sin \theta_2)$

$$\therefore OP_1 \times OP_2 = r_1 r_2 \{\cos \theta_1 \cos \theta_2 + j \sin \theta_1 \cos \theta_2 + j \cos \theta_1 \sin \theta_2 + j^2 \sin \theta_1 \sin \theta_2\}$$
$$= r_1 r_2 \{(\cos \theta_1 \cos \theta_2 - \sin \theta_1 \sin \theta_2) + j (\sin \theta_1 \cos \theta_2 + \cos \theta_1 \sin \theta_2)\}$$
$$= r_1 r_2 \{\cos (\theta_1 + \theta_2) + j \,.\, \sin (\theta_1 + \theta_2)\}$$
$$= r_1 r_2 \underline{/\theta_1 + \theta_2} \qquad (79)$$

This shows that when two vectors are multiplied, using their polar form, the moduli must be multiplied and the angles added. It can be seen that this result agrees with the fact that multiplication

by j rotates through 90°. For $j = 1\,\underline{/90°}$; hence multiplication by j will multiply the modulus by 1 and add 90° to the angle.

In general, multiplication by a vector of the form $1\,\underline{/\theta}$ is equivalent simply to a rotation through an angle θ.

When two vectors are divided, the moduli are divided and the angles subtracted.

Thus the vector
$$\frac{6\,\underline{/27°}}{3\,\underline{/11°}} = \frac{6}{3}\,\underline{/27° - 11°}$$
$$= 2\,\underline{/16°}$$

Example.—

Find the modulus and angle of $\dfrac{4 - j3}{2 + j}$

This is in the form of $\left(\dfrac{\text{vector}}{\text{vector}}\right)$. Its modulus is therefore found by dividing the moduli.

$$\therefore \text{Modulus} = \frac{|4 - j3|}{|2 + j1|} = \frac{\sqrt{4^2 + 3^2}}{\sqrt{2^2 + 1^2}} = \frac{\sqrt{16 + 9}}{\sqrt{4 + 1}} = \frac{\sqrt{25}}{\sqrt{5}} = \sqrt{5}$$

The angle is found by subtraction of the two individual angles :—

$$\begin{aligned} \text{Angle} &= \tan^{-1}\frac{-3}{4} - \tan^{-1}\frac{1}{2} \\ &= -36° 52' - 26° 34' \\ &= -63° 26' \quad \textit{Ans.} \end{aligned}$$

Raising vectors to powers

Just as multiplication of vectors can be effected in two different ways, according to whether they are expressed in rectangular or polar co-ordinates, so also can the raising of vectors to powers. When expressed in rectangular co-ordinates, a vector is raised to a power by straightforward algebraic multiplication, thus :—

$$\begin{aligned} (2 + j\,1)^2 &= (2 + j1) \times (2 + j1) \\ &= 4 + j4 - 1 \\ &= 3 + j4. \end{aligned}$$

From the rules for multiplication and division of vectors, it follows that :—

$$[r\,\underline{/\theta}]^2 = (r \times r)\,\underline{/\theta + \theta} = r^2\,\underline{/2\theta}$$

and $$[r\,\underline{/\theta}]^3 = (r \times r \times r)\,\underline{/\theta + \theta + \theta} = r^3\,\underline{/3\theta}$$

or $$[r\,\underline{/\theta}]^n = r^n\,\underline{/n\theta} \tag{80}$$

It can be seen that the treatment of vectors in polar notation is much simpler than that of vectors in rectangular notation ; since the answers to many vector calculations in electrical work are required in polar form, it is frequently advantageous to convert from rectangular to polar form before raising to powers.

Example.—

Find $(2 + j)^2$

$$(2 + j1)^2 = [\sqrt{2^2 + 1^2}, \tan^{-1} \tfrac{1}{2}]^2$$
$$= [\sqrt{5} \angle 26° \, 34']^2$$
$$= (\sqrt{5})^2 \angle 2 \times 26° \, 34'$$
$$= 5 \angle 53° \, 8' \quad \textit{Ans.}$$

As a check that this is the same result as obtained by the algebraic method above, *viz.* $(2 + j1)^2 = 3 + j4$:—

$$|3 + j4| = \sqrt{3^2 + 4^2} = \sqrt{9 + 16} = \sqrt{25} = 5$$

and angle $(3 + j4) = \tan^{-1} \dfrac{4}{3} = \tan^{-1} 1{\cdot}333 = 53° \, 8'$

Hence $5 \angle 53° \, 8' = 3 + j4$. *Ans.*

Square root of a vector

When a vector is expressed in polar form, its square root can be found by *raising the vector to the power of one-half*, by the method just described;

for example, $\sqrt{5 \angle 53° \, 8'} = \left[5 \angle 53° \, 8'\right]^{\frac{1}{2}} = \sqrt{5} \angle 26° \, 34'$.

The square root of a vector expressed in rectangular form may be determined as follows :—

Let $\sqrt{x + jy}$ be $a + jb$

Required to find a and b.

Since $a + jb = \sqrt{x + jy}$

$$\therefore \quad a^2 + 2jab - b^2 = x + jy. \tag{81}$$

Equating real quantities,

$$a^2 - b^2 = x \tag{82}$$

Equating imaginary quantities,

$$2ab = y \tag{83}$$

But :— $(a^2 + b^2)^2 = (a^2 - b^2)^2 + 4a^2b^2$

$$\therefore \quad (a^2 + b^2)^2 = x^2 + y^2$$

$$\therefore \quad a^2 + b^2 = \sqrt{x^2 + y^2} \tag{84}$$

Adding (82) and (84) :—

$$2a^2 = \sqrt{x^2 + y^2} + x$$

$$\therefore \quad a^2 = \frac{\sqrt{x^2 + y^2} + x}{2} \tag{85}$$

Subtracting (82) from (84) :—

$$2b^2 = \sqrt{x^2 + y^2} - x$$

$$b^2 = \frac{\sqrt{x^2 + y^2} - x}{2} \tag{86}$$

Thus a and b are given by (85) and (86).

Example.—

Find $\sqrt{3+j4}$.

Let $\sqrt{3+j4} = a + jb$

From (85) :—

$$a^2 = \frac{\sqrt{3^2+4^2}+3}{2}$$

$$\therefore \quad a^2 = 4$$

$$\therefore \quad a = 2$$

From (86) :—

$$b^2 = \frac{\sqrt{3^2+4^2}-3}{2}$$

$$b^2 = 1$$

$$b = 1$$

Thus $\sqrt{3+j4} = 2 + j1$ *Ans.*

As mentioned above, it is frequently advantageous to convert a vector to polar co-ordinates before manipulating it, and this example could also have been worked as follows :—

$$\begin{aligned}\sqrt{(3+j4)} &= \sqrt{\left[\sqrt{3^2+4^2}\,\underline{/\tan^{-1}\tfrac{4}{3}}\right]}\\ &= \sqrt{\left[5\,\underline{/53^\circ\,8'}\right]}\\ &= \left[5\,\underline{/53^\circ\,8'}\right]^{\frac{1}{2}}\\ &= \sqrt{5}\,\underline{/26^\circ\,34'} \quad \textit{Ans.}\end{aligned}$$

Rationalisation

It is sometimes necessary to eliminate j from the denominator of an expression. This can be done by " rationalisation ", *i.e.*, by multiplication of both numerator and denominator by the " conjugate " of the denominator.

The " *conjugate* " of a vector $r\,\underline{/\theta} = x + jy$ is that vector which has the same modulus and equal but opposite angle. Thus, the conjugate of $r\,\underline{/\theta}$ is $r\,\underline{/-\theta}$ and the conjugate of $x + jy$ is $x - jy$ (87)

Example.—

Express $\dfrac{1}{5-j3}$ in the rectangular form $x + jy$

$$\frac{1}{5-j3} = \frac{1}{5-j3} \times \frac{5+j3}{5+j3} = \frac{5+j3}{5^2-j^2 3^2} = \frac{5+j3}{25+9} = \frac{5}{34} + j\frac{3}{34} \quad \textit{Ans.}$$

It is a good rule to avoid rationalisation unless necessary.

Exponential notation of a vector

As will be seen later (equation 125 p. 91), $\cos\theta + j\sin\theta = e^{j\theta}$. Thus the vector $r\,\underline{/\theta} \equiv r(\cos\theta + j\sin\theta)$ may be written as the vector $r\,.\,e^{j\theta}$.

A further apparent simplification can be carried out by letting $r = \epsilon^a$. The vector now becomes $\epsilon^a \cdot \epsilon^{j\theta} = \epsilon^{a+j\theta}$.

By letting $a + j\theta$ equal the complex number γ say, the vector $r\angle\theta$ may be denoted simply and completely as the vector ϵ^γ. Thus —

$$OP = r\angle\theta = \epsilon^{a+j\theta} = \epsilon^\gamma \qquad (88)$$

Logarithm of a vector

To find $\log_e[r\angle\theta]$.

Rewrite the vector $[r\angle\theta]$ using the exponential notation as —

$$[r\angle\theta] = r\,\epsilon^{j\theta}$$

Then
$$\log_e[r\angle\theta] = \log_e(r\,\epsilon^{j\theta})$$
$$= \log_e r + \log_e \epsilon^{j\theta}$$
$$= \log_e r + j\theta \qquad (89)$$

Rotating vectors

Let the vector OP of length r rotate in an anti-clockwise direction about a centre O. Then when OP has turned through an angle θ from its horizontal position along Ox (see Fig. 53a) the instantaneous height h of P above Ox is given by

$$h = r \sin\theta \qquad (90)$$

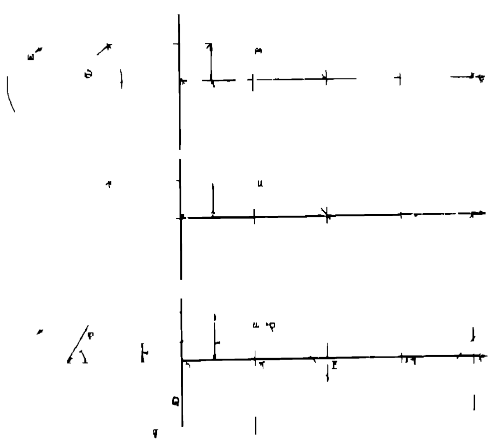


Fig. 53.—Rotating vectors.

If the vertical height h of P be plotted against θ the resulting curve is as shown on the right of Fig. 53a. Since the one variable (h) is directly proportional to the *sine* of the other (θ) this is called a "sine," or "sinusoidal" curve.

If OP be rotating with a constant angular velocity ω corresponding to f revolutions per second—that is, $\omega = 2\pi f$ radians per second—then the time for one revolution is $\frac{1}{f} = \frac{2\pi}{\omega}$ seconds. After time t, OP will have turned through an angle $\theta = \omega t$, where t is measured from the instant when OP is in the horizontal position along Ox. The vertical height of P at any time t is given by :—

$$h = r \sin \omega t \tag{91}$$

and varies as shown in Fig. 53*b*. Time may, however, be measured from some instant other than that at which OP lies along Ox. Consider a vector OQ that makes an angle φ with OP; then at the instant $t = 0$, OQ makes an angle φ with Ox, and at any later instant t, OQ will make an angle $(\omega t + \varphi)$ with Ox. The height k of Q at time t is then given by :—

$$k = r \sin (\omega t + \varphi) \tag{92}$$

This is shown in Fig. 53*c*.

Many mechanical and electrical quantities vary with time in a sinusoidal manner, and are represented by equations of the form (92) and by figures similar to Fig. 53. Rotating vectors provide a very convenient method of expressing such sinusoidal waveforms; the addition, subtraction, multiplication and division of sine waves can then be easily achieved by the application of these processes to the rotating vectors, whereas such operations can not so easily be applied to the sine waves themselves.

Addition of two sine waves using rotating vectors

Consider the two sinusoidal waveforms represented by the equations :

$$a = A \sin \omega t$$
$$b = B \sin (\omega t + \varphi)$$

where a and b represent the instantaneous heights of the two curves at any time t, and A and B are the peak or maximum heights. These two sine waves are illustrated in Fig. 54 *a* and *b* respectively, and they could clearly be added, to give curve (*c*), by the somewhat tedious process of adding the individual heights of the two curves at each instant.

A much easier way of adding the two curves is to draw two vectors $OP = A \,\underline{/\omega t}$ and $OQ = B \,\underline{/\omega t + \varphi}$, such that, if they be rotated at constant angular velocity ω, the graphs of the instantaneous heights of P and Q will be identical with the curves to be added. These two vectors (Fig. 55) then represent the two curves, and they may be added by completing the parallelogram. The result (Fig. 56) is a third vector OR which can be denoted by $C \,\underline{/\omega t + \theta}$, where C and θ are, as shown below, given by :—

$$C = |\,OR\,| = \sqrt{A^2 + B^2 + 2AB \cos \varphi} \tag{93}$$

$$\text{and} \qquad \theta = \tan^{-1}\left(\frac{B \sin \varphi}{A + B \cos \varphi}\right) \tag{94}$$

This vector OR is the sum of the two vectors OP and OQ, and represents curve $c = C \sin(\omega t + \theta)$ which is the sum of the curves represented by OP and OQ

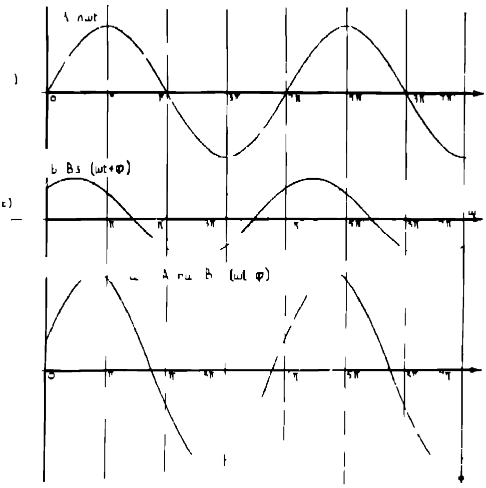


Fig 54 —Addition of two sine waves

The values of the modulus C and angle $\angle\theta$ of the resultant vector $OR = OP + OQ$ can be evaluated as follows

From Fig 55

$OP = A \cos \omega t + jA \sin \omega t$

$OQ = B \cos(\omega t + \varphi) + jB \sin(\omega t + \varphi)$

$OR = OP + OQ$

$\therefore \quad OR = \{A \cos \omega t + B \cos(\omega t + \varphi)\} + j\{A \sin \omega t + B \sin(\omega t + \varphi)\}$

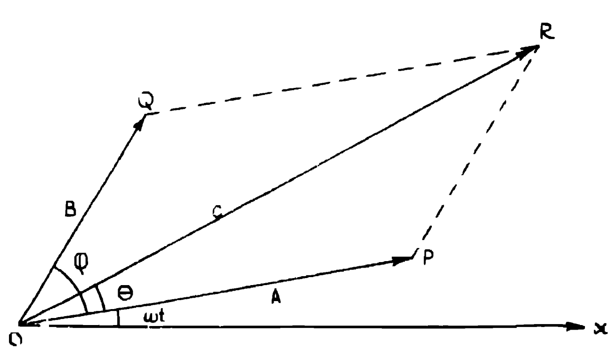


Fig 55 —Addition of two vectors OP and OQ representing two sine waves

$$\therefore |OR|^2 = \{A\cos\omega t + B\cos(\omega t + \varphi)\}^2 + \{A\sin\omega t + B\sin(\omega t + \varphi)\}^2$$
$$= \{A^2\cos^2\omega t + B^2\cos^2(\omega t + \varphi) + 2AB\cos(\omega t + \varphi)\cos\omega t\}$$
$$+ \{A^2\sin^2\omega t + B^2\sin^2(\omega t + \varphi) + 2AB\sin(\omega t + \varphi)\sin\omega t\}$$
$$= A^2 + B^2 + 2AB\{\cos(\omega t + \varphi)\cos\omega t + \sin(\omega t + \varphi)\sin\omega t\}$$
$$= A^2 + B^2 + 2AB\cos\Big\{(\omega t + \varphi) - (\omega t)\Big\}$$
$$= A^2 + B^2 + 2AB\cos\varphi$$
$$\therefore \quad C = |OR| = \sqrt{A^2 + B^2 + 2AB\cos\varphi} \qquad (95)$$

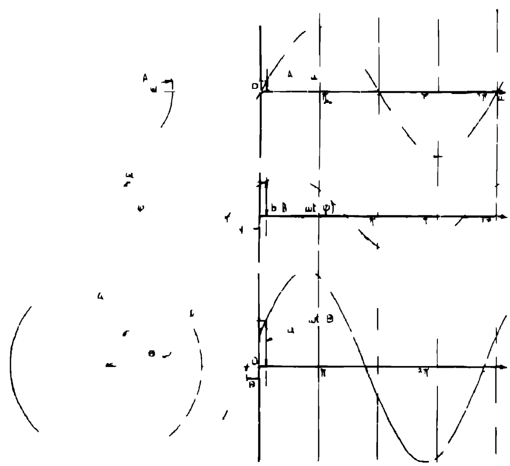


Fig 56—Addition of two sine waves by means of rotating vectors

The angle made by OR with Ox at any instant t is $(\omega t + \theta)$, so that from Fig 56 —

$$\tan(\omega t + \theta) = \frac{A\sin\omega t + B\sin(\omega t + \varphi)}{A\cos\omega t + B\cos(\omega t + \varphi)}$$

θ can be expressed as the difference between two angles i.e. —

$$\tan\theta = \tan\Big\{(\omega t + \theta) - \omega t\Big\}$$

so that

$$\tan\theta = \frac{\tan(\omega t + \theta) - \tan\omega t}{1 + \tan(\omega t + \theta)\tan\omega t}$$

$$\therefore \tan\theta = \frac{\dfrac{A\sin\omega t + B\sin(\omega t + \varphi)}{A\cos\omega t + B\cos(\omega t + \varphi)} - \dfrac{\sin\omega t}{\cos\omega t}}{1 + \dfrac{A\sin\omega t + B\sin(\omega t + \varphi)}{A\cos\omega t + B\cos(\omega t + \varphi)} \cdot \dfrac{\sin\omega t}{\cos\omega t}}$$

$$= \frac{B\sin\{(\omega t + \varphi) - \omega t\}}{A(\sin^2\omega t + \cos^2\omega t) + B\cos\Big\{(\omega t + \varphi) - \omega t\Big\}}$$

$$= \frac{B\sin\varphi}{A + B\cos\varphi} \qquad (96)$$

PART II

MORE ADVANCED MATHEMATICS

DIFFERENTIAL CALCULUS

Differential calculus deals with the *rate of change* of a function: *e.g.*, if $y = f(x)$, by how much will y increase (or decrease) if the value of x is increased by some small amount? For functions in general, the answer will of course depend on the initial value of x. The graphical illustration of this is useful: if the curve $y = f(x)$ is drawn, the rate of change of y will be proportional to the *slope* of the curve. If it is very steep, a small increase in x will produce a proportionately large increase in y; if the curve is fairly flat, the increase in y will be small. It can be seen that for any curve except a straight line the slope, or rate of change, varies from place to place. Hence, if it is to be measured by considering the ratios of increases in y and x, it is important to see that these increases are small, otherwise their ratio will not give a true idea of the slope of the curve at that point.

These small changes in the values of x and y are usually denoted by δx or δy (alternatively Δx or Δy)—the Greek letter delta signifying "a small change in . . . .".

Note that if δx is small, $(\delta x)^2$ will be even smaller, and can be neglected in comparison with δx. Higher powers of δx are, of course, smaller still.

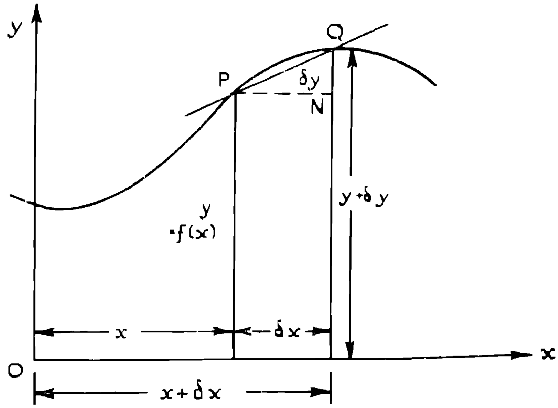


Fig. 57.—Differential notation.

It has been stated that the rate of change, or slope, is given by the ratio of the increases in y and x, or $\frac{\delta y}{\delta x}$ as it would be written. This is illustrated graphically in Fig. 57.

To find the slope at $P \equiv (x, y)$, an adjacent point Q is taken, whose co-ordinates are $(x + \delta x, y + \delta y)$. The ratio $\frac{\delta y}{\delta x}$ will give the tangent of the angle QPN, *i.e.*, the "slope" of the line PQ. It can be seen that if Q is made very close to P, this will give the slope of the curve at P. The conditions for this are that δy and δx both become zero. Although they both become zero, as Q moves up to P, the ratio $\frac{\delta y}{\delta x}$ approaches some finite value, which will be the slope of the curve. This value is denoted by $\frac{dy}{dx}$. In other words, $\frac{dy}{dx}$ is the limit of the ratio $\frac{\delta y}{\delta x}$ as δx (and therefore δy) $\to 0$, or :—

$$\frac{dy}{dx} = \lim_{\delta x \to 0} \frac{\delta y}{\delta x} \tag{97}$$

Hence $\frac{dy}{dx}$ gives the rate of change of y, or the slope, at this point (x, y). As the slope depends upon the position of the point P, it is obvious that $\frac{dy}{dx}$ must be a function of x. To find the slope at any *particular* point the numerical value of x must be inserted.

Notation

y is a function of x, denoted by $f(x)$. $\frac{dy}{dx}$ is a function of x, and is denoted by $f'(x)$; this is called the "first derivative", or "differential coefficient", of $f(x)$, *i.e.* :—

$$\frac{dy}{dx} = f'(x) \tag{98}$$

This is sometimes written as $\frac{d}{dx}f(x)$

Calculation of $f'(x)$ ("differentiation")

$f'(x)$ has been defined as the limit of $\frac{\delta y}{\delta x}$ when δx tends to zero. This indicates the method of calculation.

At P, $\quad y = f(x) \quad$ (i)

At Q, x increases to $x + \delta x$, and as a result, y increases to $y + \delta y$.

$\therefore \quad y + \delta y = f(x + \delta x) \quad$ (ii)

This is the condition that Q shall lie on the curve.

Subtracting equation (i) from equation (ii) :—

$$\delta y = f(x + \delta x) - f(x)$$

Hence
$$\frac{\delta y}{\delta x} = \frac{f(x + \delta x) - f(x)}{\delta x}$$

Thus
$$f'(x) = \frac{dy}{dx}$$
$$= \underset{\delta x \to 0}{\text{Limit}} \frac{\delta y}{\delta x}$$
$$= \underset{\delta x \to 0}{\text{Limit}} \frac{f(x + \delta x) - f(x)}{\delta x} \qquad (99)$$

This is the basic formula from which all derivatives are calculated.

Example :—

Take a particular case, say $y = x^2$, and find the slope of this curve at the point (x, y).

$$f(x) = x^2$$

From above,
$$\frac{dy}{dx} = \underset{\delta x \to 0}{\text{Limit}} \frac{f(x + \delta x) - f(x)}{\delta x}$$

Now
$$\frac{\delta y}{\delta x} = \frac{(x + \delta x)^2 - x^2}{\delta x}$$
$$= \frac{x^2 + 2x \cdot \delta x + \delta x^2 - x^2}{\delta x}$$
$$= 2x + \delta x$$

Note that no approximations have been made

But $\frac{dy}{dx}$ is the limit of $\frac{\delta y}{\delta x}$ when $\delta x \to 0$
and equals the limit of $2x + \delta x$ when $\delta x \to 0$

$$\therefore \qquad \frac{dy}{dx} = 2x$$

Thus the slope of $y = x^2$ at any point is equal to $2x$ (*e g*, at the point (3, 9), the slope is $2 \times 3 = 6$)

Derivative of x^n

Let
$$f(x) = x^n$$

Then
$$\frac{\delta y}{\delta x} = \frac{(x + \delta x)^n - x^n}{\delta x}$$

That is.—

$$\frac{\delta y}{\delta x} = \frac{x^n\left(1 + \frac{\delta x}{x}\right)^n - x^n}{\delta x}$$

Expand $\left(1 + \frac{\delta x}{x}\right)^n$ by the Binomial theorem, where n and x may have any value, since $\left|\frac{\delta x}{x}\right| < 1$

$$\frac{\delta y}{\delta x} = \frac{x^n\left\{1 + n \cdot \frac{\delta x}{x} + \frac{n(n-1)}{\underline{|2}}\left(\frac{\delta x}{x}\right)^2 + \quad \ldots \right\} - x^n}{\delta x}$$

$$\therefore \qquad \frac{\delta y}{\delta x} = \frac{\left\{x^n + n\delta x \; x^{n-1} + \frac{n(n-1)}{\underline{|2}}(\delta x)^2 x^{n-2} + \ldots\right\} - x^n}{\delta x}$$

$$= nx^{n-1} + \frac{n(n-1)}{\underline{|2}} x^{n-2}\,\delta x + \ldots$$

All terms after the second involve powers of δx greater than the first, and may be neglected when $\delta x \to 0$

$$\therefore \qquad \frac{dy}{dx} = \underset{\delta x \to 0}{\text{Limit}} \frac{\delta y}{\delta x} = nx^{n-1}$$

i.e.,
$$\frac{d}{dx}(x^n) = nx^{n-1} \tag{100}$$

e.g., as already proved, $\frac{d}{dx}x^2 = 2x$.

Example.—

Differentiate $\frac{1}{\sqrt{x}}$ with respect to x

Let
$$y = \frac{1}{\sqrt{x}} = x^{-\frac{1}{2}}$$

$$\therefore \qquad \frac{dy}{dx} = -\tfrac{1}{2}x^{-\frac{3}{2}} \quad \textit{Ans.}$$

Constants

The equation $y = C$, where C is a constant, represents a straight line parallel to the x-axis. The slope is zero, so that :—

If
$$y = C$$

Then
$$\frac{dy}{dx} = \frac{d}{dx}(C) = 0 \tag{101}$$

Multiplication by a constant.—If the function is *multiplied* by some constant, the constant remains after differentiation, for example, the slope of the curve $y = 5x^3$ is five times as great as that of $y = x^3$.

Thus if
$$y = 5x^3$$
$$\frac{dy}{dx} = 5 \times 3x^2 = 15x^2$$

Products

Consider $y = f(x) = u \, . \, v$, where u and v are both functions of x.

Then
$$\frac{\delta y}{\delta x} = \frac{(u + \delta u)(v + \delta v) - uv}{\delta x}$$

$$= \frac{uv + u \, . \, \delta v + v \, . \, \delta u + \delta u \, . \, \delta v - uv}{\delta x}$$

$$= u \, . \frac{\delta v}{\delta x} + v \, . \frac{\delta u}{\delta x} + \frac{\delta u \, . \, \delta v}{\delta x}$$

As δx tends to zero, so also will δu and δv. The term $\frac{\delta u \, . \, \delta v}{\delta x}$ will then be small and may be neglected.

$$\frac{dy}{dx} = \underset{\delta x \to 0}{\text{Limit}} \frac{\delta y}{\delta x} = u \frac{dv}{dx} + v \frac{du}{dx}$$

i.e.,
$$\frac{d}{dx}(u \cdot v) = u \frac{dv}{dx} + v \frac{du}{dx} \tag{102}$$

Example —

Find $\frac{dy}{dx}$ when $y = (x^2 + 3)(2x - 1) = u \cdot v$,

where $u = x^2 + 3$ and $v = 2x - 1$

$$\frac{dy}{dx} = (x^2 + 3) \cdot 2 + 2x \cdot (2x - 1)$$

$$= 2x^2 + 6 + 4x^2 - 2x$$

$$\therefore \frac{dy}{dx} = 6x^2 - 2x + 6 \quad \textit{Ans}$$

Quotients

In a similar manner to the above, it may be shown that if $y = \frac{u}{v}$, where u and v are both functions of x, then —

$$\frac{dy}{dx} = \frac{d}{dx}\left(\frac{u}{v}\right) = \frac{v\frac{du}{dx} - u\frac{dv}{dx}}{v^2} \tag{103}$$

Example —

Find $\frac{dy}{dx}$ when $y = \frac{x^2 - 1}{x + 2}$

In this case $u = x^2 - 1$ and $v = x + 2$

$$\frac{dy}{dx} = \frac{(x + 2) \cdot 2x - (x^2 - 1) \cdot 1}{(x + 2)^2}$$

$$\therefore \frac{dy}{dx} = \frac{x^2 + 4x + 1}{(x + 2)^2} \quad \textit{Ans}$$

Trigonometrical functions

To take another particular case let $y = \sin x$. It is required to find the slope of this curve at the point (x, y)

$$f(x) = \sin x$$

$$\frac{dy}{dx} = \underset{\delta x \to 0}{\text{Limit}} \frac{f(x + \delta x) - f(x)}{\delta x}$$

$$= \underset{\delta x \to 0}{\text{Limit}} \frac{\sin (x + \delta x) - \sin x}{\delta x}$$

$$= \underset{\delta x \to 0}{\text{Limit}} \frac{\sin x \cos \delta x + \cos x \sin \delta x - \sin x}{\delta x}$$

When $\delta x \to 0$, $\cos \delta x \to 1$ and $\sin \delta x \to \delta x$

$$\frac{dy}{dx} = \frac{\sin x \cdot 1 + \cos x \cdot \delta x}{\delta x} - \frac{\sin x}{\delta x} = \cos x$$

Thus
$$\frac{d}{dx}(\sin x) = \cos x \tag{104}$$

This result may be verified by reference to Fig. 58.

When $x = 0$, $y = \sin x = 0$, $\cos x = 1$, and the slope of $y = \sin x$ is a maximum.

When $x = 90°$, $y = \sin x = 1$, $\cos x = 0$, and the slope of $y = \sin x$ is also zero.

When $x = 180°$, $y = \sin x = 0$, $\cos x = -1$, and the slope of $y = \sin x$ is a maximum in a negative direction.

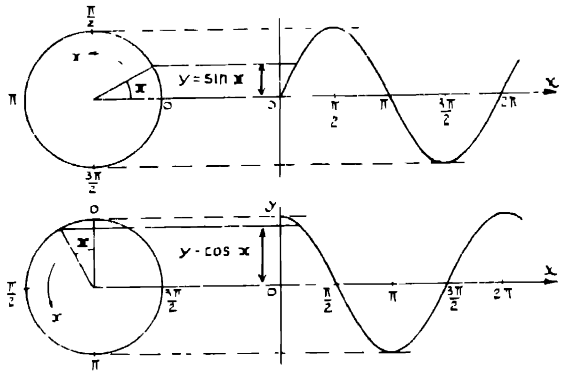


FIG. 58.—Graphs of $y = \sin x$ and $y = \frac{d}{dx} \sin x = \cos x$.

It has been shown that the differential coefficient of $\sin x$ is $\cos x$. In a similar manner it may be shown that the differential coefficient of $\cos x$ is $-\sin x$.

Consider the differential coefficient of $\tan x$.

Let $y = \tan x$

$\therefore$ $y = \dfrac{\sin x}{\cos x}$

This may be considered as a quotient. Hence :—

$$\frac{dy}{dx} = \frac{\cos x \dfrac{d(\sin x)}{dx} - \sin x \dfrac{d(\cos x)}{dx}}{\cos^2 x}$$

$$= \frac{\cos^2 x + \sin^2 x}{\cos^2 x}$$

$$= \frac{1}{\cos^2 x} = \sec^2 x$$

Thus $\dfrac{d}{dx}(\tan x) = \sec^2 x$ (105)

The derivatives of other trigonometrical functions may be calculated in a similar manner. Some of the more important ones

are listed in Table II; a further list of derivatives is given in Appendix I.

TABLE II

Differential coefficients of the trigonometrical ratios

y	$\frac{dy}{dx}$
$\sin x$	$\cos x$
$\cos x$	$-\sin x$
$\tan x$	$\sec^2 x$
$\cot x$	$-\operatorname{cosec}^2 x$
$\sec x$	$\sec x \,.\, \tan x$
$\operatorname{cosec} x$	$-\operatorname{cosec} x \,.\, \cot x$

Inverse functions

The differentiation of an inverse function (such as $\sin^{-1} x$ or $\tan^{-1} x$) is interesting. Consider

$$y = \sin^{-1} x$$

$\frac{dy}{dx}$ cannot be found at once; rewrite the equation as :—

$$x = \sin y$$

Then
$$\frac{dx}{dy} = \cos y$$

$$\frac{dy}{dx} = \frac{1}{\cos y}$$

Now
$$\cos y = \sqrt{1 - \sin^2 y} = \sqrt{1 - x^2}$$

$$\frac{dy}{dx} = \frac{1}{\sqrt{1 - x^2}}$$

Similarly, if $y = \tan^{-1} x$

$$x = \tan y$$

$$\frac{dx}{dy} = \sec^2 y$$

$$\frac{dy}{dx} = \frac{1}{\sec^2 y} = \frac{1}{1 + \tan^2 y} = \frac{1}{1 + x^2}$$

Thus
$$\frac{d}{dx}(\tan^{-1} x) = \frac{1}{1 + x^2} \qquad (106)$$

Sum of terms and series

If $y = f(x)$ is given as a sum of a number of terms, $\frac{dy}{dx}$ is found by adding the derivatives of each term taken individually.

Thus, if
$$y = 5x^2 + 3x + 7$$
$$\frac{dy}{dx} = (5 \times 2x) + 3 + 0$$
$$= 10x + 3$$

Exponential and logarithmic functions

The derivative of an exponential function may be found by expressing it as a series; each term of the series can then be differentiated, and the resulting derivatives added.

Thus if :— $y = e^x$

It can be expressed as :—
$$y = 1 + x + \frac{x^2}{\lfloor 2} + \frac{x^3}{\lfloor 3} + \frac{x^4}{\lfloor 4} + \ldots$$
$$\therefore \quad \frac{dy}{dx} = 0 + 1 + \frac{2x}{\lfloor 2} + \frac{3x^2}{\lfloor 3} + \frac{4x^3}{\lfloor 4} + \ldots$$
$$= 0 + 1 + x + \frac{x^2}{\lfloor 2} + \frac{x^3}{\lfloor 3} + \ldots$$
$$= e^x$$

Thus
$$\frac{d}{dx}(e^x) = e^x \tag{107}$$

Using this result, one can find the derivative of $\log_e x$.

For, if $y = \log_e x$

Then :— $x = e^y$
$$\therefore \quad \frac{dx}{dy} = \frac{d}{dy}(e^y)$$
$$= e^y$$
$$\therefore \quad \frac{dy}{dx} = \frac{1}{e^y}$$
$$= \frac{1}{x}$$

This gives the important relationship that :—
$$\frac{d}{dx}(\log_e x) = \frac{1}{x} \tag{108}$$

Function of a function

$$y = F[f(x)]$$

Each function is differentiated and the results multiplied :—

$$\frac{dy}{dx} = F'[f(x)] \times f'(x) \tag{109}$$

This can be extended for functions of a function of a function.

Example 1.—

Differentiate $\sin x^2$ with respect to x.

This is a function (sin) of a function of x, (x^2).

First differentiate $\sin x^2$ with respect to x^2, giving $\cos x^2$.

Next differentiate x^2 with respect to x, giving $2x$.

Then the differential coefficient of $\sin x^2$ with respect to x is :—

$$\cos x^2 \times 2x = 2x \cdot \cos x^2 \quad \textit{Ans.}$$

Example 2.—

Find $\frac{dy}{dx}$ when $y = \sqrt{x^3 - 3x}$

Differentiate the $\sqrt{\ }$, giving $\frac{1}{2}(x^3 - 3x)^{-\frac{1}{2}}$

Differentiate $x^3 - 3x$, giving $3x^2 - 3$

$$\therefore \quad \frac{dy}{dx} = \frac{3}{2}(x^2 - 1)(x^3 - 3x)^{-\frac{1}{2}} \quad \textit{Ans.}$$

Example 3.—

$$y = e^{3x^2}$$

$$\frac{dy}{dx} = e^{3x^2} \cdot \frac{d}{dx}(3x^2)$$

$$= 6x \cdot e^{3x^2} \quad \textit{Ans.}$$

Example 4.—

$$y = \log \tan \sqrt{x^2 + 1}. \text{ Find } \frac{dy}{dx}.$$

This is a function (log) of a function (tan) of a function ($\sqrt{\ }$) of the function $(x^2 + 1)$, which is a function of x.

Differentiating $\log \tan \sqrt{x^2 + 1}$ with respect to $\tan \sqrt{x^2 + 1}$

gives :— $\dfrac{1}{\tan \sqrt{x^2 + 1}}$

Differentiating $\tan \sqrt{x^2 + 1}$ with respect to $\sqrt{x^2 + 1}$

gives :— $\sec^2 \sqrt{x^2 + 1}$

Differentiating $\sqrt{(x^2 + 1)}$ with respect to $(x^2 + 1)$

gives :— $\frac{1}{2}(x^2 + 1)^{-\frac{1}{2}}$

Differentiating $(x^2 + 1)$ with respect to x

gives :— $2x$.

$$\text{Hence } \frac{d}{dx}(\log \tan \sqrt{x^2 + 1}) = \frac{1}{\tan \sqrt{x^2 + 1}} \times \sec^2 \sqrt{x^2 + 1} \times \frac{1}{2}(x^2 + 1)^{-\frac{1}{2}} \times 2x$$

$$= \frac{x}{\sqrt{x^2 + 1} \sin \sqrt{x^2 + 1} \cos \sqrt{x^2 + 1}} \quad \textit{Ans.}$$

Successive differentiation

If $\frac{dy}{dx}$ is itself differentiated with respect to x, the result

written $\frac{d^2y}{dx^2}$ or $f''(x)$. This gives the rate of change of the slope of $f(x)$, and is known as the *second* derivative or differential coefficient. Similarly subsequent derivatives may be found; their calculation involves no new principles.

Example.—

Find the third derivative of $y = x^3 \sin x$.

$$\frac{dy}{dx} = 3x^2 \,.\, \sin x + x^3 \cos x$$

$$\frac{d^2y}{dx^2} = 6x \sin x + 3x^2 \cos x + 3x^2 \cos x - x^3 \sin x$$

$$= 6x \sin x + 6x^2 \cos x - x^3 \sin x$$

$$\frac{d^3y}{dx^3} = 6 \sin x + 6x \cos x + 12x \cos x - 6x^2 \sin x - x^3 \cos x - 3x^2 \sin x$$

$$= 6 \sin x + 18x \cos x - 9x^2 \sin x - x^3 \cos x. \quad \textit{Ans.}$$

Maxima and Minima

Since $\frac{dy}{dx}$ gives the slope of $f(x)$ at any point, it can be used to find the points where y is a maximum or minimum; for at these points the curve is horizontal, *i.e.* the slope is zero. The values of x that make y a maximum or minimum will therefore be the roots of the equation $f'(x) = 0$.

To decide whether any particular root of this equation gives a maximum or minimum, two methods may be used.

1.—Find $\frac{dy}{dx}$ for values of x *just* above and below that giving $f'(x) = 0$. If it is positive below and negative above, the curve will be as Fig. 59, *i.e.* a maximum. If it is negative below and positive above, the point will be a minimum.

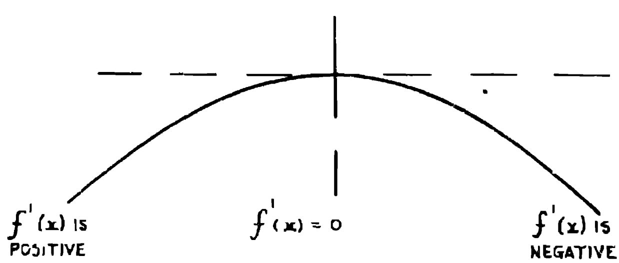


FIG. 59.—Curve with maximum.

It is possible sometimes for the slope to have the same sign before and after. In this case the curve is as shown in Fig. 60, and the point is neither a maximum nor a minimum, but is a "point of inflection".

2.—For a maximum, the slope is positive before and negative after—*i.e.*, the slope is decreasing, therefore the rate of change of slope is negative. $\frac{d^2y}{dx^2}$ is therefore negative. Similarly at a minimum $\frac{d^2y}{dx^2}$ is positive. Hence an alternative way is to find $\frac{d^2y}{dx^2}$ at the point; if it is negative the point is a maximum, if positive it is a minimum. If $\frac{d^2y}{dx^2} = 0$ the curve may be of the form shown in Fig. 60.

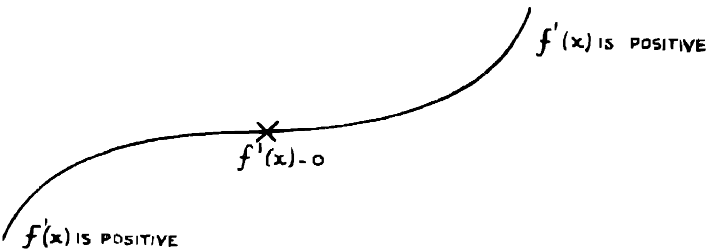


FIG. 60.—Curve with point of inflection.

Example 1.—

Consider the curve:—

$$y = x^4 - 5x^2 + 4$$

$$\frac{dy}{dx} = 4x^3 - 10x$$

Maxima or minima occurs if $4x^3 - 10x = 0$

i.e., if $x = 0$ or $\pm\sqrt{\frac{5}{2}}$

Now determine whether these points are maxima or minima.

Method 1.—

Consider the point $x = 0$; just before $\frac{dy}{dx}$ is positive, just after it is negative.

Therefore $x = 0$ is a maximum.

Consider the point $x = +\sqrt{\frac{5}{2}}$, just before $\frac{dy}{dx}$ is negative, just after it is positive.

Therefore $x = +\sqrt{\frac{5}{2}}$ is a minimum.

Consider the point $x = -\sqrt{\frac{5}{2}}$; just before, $\frac{dy}{dx}$ is negative, just after, it is positive.

Therefore $x = -\sqrt{\frac{5}{2}}$ is also a minimum

Method 2.—

$$\frac{d^2y}{dx^2} = 12x^2 - 10$$

At $x = 0$ this is negative, therefore at $x = 0$ the curve has a maximum

At $x = \pm\sqrt{\frac{5}{2}}$ this is positive, therefore at these points the curve has minima

Thus the curve has a maximum at $x = 0$ and minima at $x = +\sqrt{\frac{5}{2}}$ and at $x = -\sqrt{\frac{5}{2}}$

The curve is shown in Fig 61

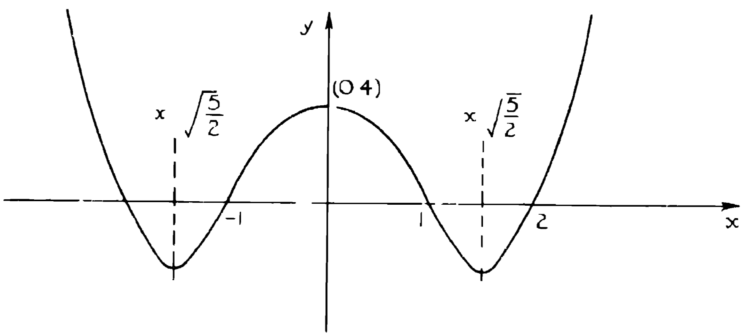


Fig 61 —Graph of $y = x^4 - 5x^2 + 4$

Example 2 —

What positive value of L will make

$$\sqrt{(R^2 + \omega^2L^2)(G^2 + \omega^2C^2)} + RG - \omega^2LC$$

a maximum or minimum?

Putting $\quad y = \sqrt{(R^2 + \omega^2L^2)(G^2 + \omega^2C^2)} + RG - \omega^2LC$

$$\frac{dy}{dL} = \frac{1}{2}\sqrt{\frac{G^2 + \omega^2C^2}{R^2 + \omega^2L^2}}\,2\omega^2L - \omega^2C$$

For y to be a maximum or minimum $\dfrac{dy}{dL} = 0$

$$\therefore \qquad \sqrt{\frac{G^2 + \omega^2C^2}{R^2 + \omega^2L^2}} = \frac{C}{L}$$

i e.,

$$\frac{G^2 + \omega^2C^2}{R^2 + \omega^2L^2} = \frac{C^2}{L^2}$$

i.e.,
$$L^2G^2 = R^2C^2$$
or
$$L = \frac{RC}{G}$$

This gives the value of L for y to be a maximum or minimum. It can be shown by differentiating again that this value of L gives a *minimum* value to y.

Differentiation of a vector

Vectors and complex numbers are differentiated in the same manner as other functions. If, however, the vector is given in polar form as $r\angle\theta$ it must first be written in rectangular co-ordinate or in exponential form, *i.e.*, as $r\cos\theta + jr\sin\theta$ or as $re^{j\theta}$.

Consider the differential coefficient, with respect to t, of the rotating vector

$$OP = r\angle\underline{\omega t + \varphi} \tag{110}$$

(*a*) This vector may be expressed in rectangular co-ordinates as:—

$$OP = r\cos(\omega t + \varphi) + jr\sin(\omega t + \varphi) \tag{111}$$

Then
$$\frac{d}{dt}(OP) = \frac{d}{dt}\left\{r\cos(\omega t + \varphi)\right\} + j\frac{d}{dt}\left\{r\sin(\omega t + \varphi)\right\}$$
$$= r\frac{d}{dt}\cos(\omega t + \varphi) + \cos(\omega t + \varphi)\frac{dr}{dt}$$
$$+ jr\frac{d}{dt}\sin(\omega t + \varphi) + j\sin(\omega t + \varphi)\frac{dr}{dt}$$

In the case under consideration $\frac{dr}{dt} = 0$, that is, the amplitude of the vector is not changing. Therefore

$$\frac{d}{dt}(OP) = -\omega r\sin(\omega t + \varphi) + 0 + j\omega r\cos(\omega t + \varphi) + 0$$
$$= j\omega\left\{r\cos(\omega t + \varphi) + jr\sin(\omega t + \varphi)\right\}$$
$$= j\omega\, OP$$

(*b*) Alternatively the vector OP may be expressed in exponential form as

$$OP = re^{j(\omega t + \varphi)} \tag{112}$$
$$= re^{j\omega t}e^{j\varphi}$$
$$= (e^{j\omega t})(re^{j\varphi})$$

Then
$$\frac{d}{dt}(OP) = (re^{j\varphi})\frac{d}{dt}e^{j\omega t} + e^{j\omega t}\frac{d}{dt}(re^{j\varphi})$$
$$= re^{j\varphi}\,j\omega\,e^{j\omega t} + 0$$
$$= j\omega r\,e^{j\omega t}e^{j\varphi}$$
$$= j\omega\,OP, \text{ as before}$$

Partial differentiation

Frequently a dependent variable is a function of two or more independent variables, for example, the volume V of a cylinder is determined by both the radius r and the length h, and one can write —

$$V = f(r, h) \tag{113}$$

In order to determine the rate at which V increases or decreases when one of the independent variables changes the other independent variable must be kept constant throughout the calculation. Thus while determining the rate at which the volume of a cylinder varies with radius the length h must be kept constant. A derivative of a variable with respect to one of several independent variables is known as a "partial derivative" and is represented by the symbol '∂' to distinguish it from the 'd' in the ordinary derivative $\frac{dy}{dx}$. Then —

$\frac{\partial V}{\partial r}$ the rate at which V increases with respect to r *when all other relevant independent variables (e.g. h) are kept constant.*

$\frac{\partial V}{\partial h}$ the rate at which V increases with respect to h *when all other relevant independent variables (e.g. r) are kept constant.*

Partial derivatives are evaluated in the normal manner, all variables other than the two concerned being treated as constants. Thus for example

$$\text{if} \quad V = \pi r^2 h \tag{114}$$

$$\text{Then} \quad \frac{\partial V}{\partial r} = 2\pi r h \tag{115}$$

$$\text{and} \quad \frac{\partial V}{\partial h} = \pi r^2 \tag{116}$$

If a small change δh is made in the value of h whilst r is kept constant, the corresponding change δV is given by

$$\delta V \simeq \frac{\partial V}{\partial h}\,\delta h$$

Similarly for a change δr in the value of r, h being kept constant

$$\delta V \simeq \frac{\partial V}{\partial r}\,\delta r$$

In general if r and h are changed simultaneously, then —

$$\delta V \simeq \frac{\partial V}{\partial r}\,\delta r + \frac{\partial V}{\partial h}\,\delta h$$

Second-order partial derivatives may be evaluated as follows. In the case of a function of two independent variables, *e.g.*, $V = f(r, h)$,– there are four second-order partial derivatives, *viz*

$\frac{\partial^2 V}{\partial r^2}$ = the rate at which $\frac{\partial V}{\partial r}$ increases with respect to r when h is kept constant.

$\dfrac{\partial^2 V}{\partial h^2}$ = the rate at which $\dfrac{\partial V}{\partial h}$ increases with respect to h when r is kept constant

$\dfrac{\partial^2 V}{\partial r\,\partial h}$ = the rate at which $\dfrac{\partial V}{\partial h}$ increases with respect to r when h is kept constant

$\dfrac{\partial^2 V}{\partial h\,\partial r}$ = the rate at which $\dfrac{\partial V}{\partial r}$ increases with respect to h when r is kept constant

Taking as an example the function $V = \pi r^2 h$, the second-order partial derivatives are found as follows

$$\frac{\partial^2 V}{\partial r^2} = \frac{\partial}{\partial r}\left(\frac{\partial V}{\partial r}\right) = \frac{\partial}{\partial r}(2\pi rh) = 2\pi h$$

$$\frac{\partial^2 V}{\partial h^2} = \frac{\partial}{\partial h}\left(\frac{\partial V}{\partial h}\right) = \frac{\partial}{\partial h}(\pi r^2) = 0$$

$$\frac{\partial^2 V}{\partial r\,\partial h} = \frac{\partial}{\partial r}\left(\frac{\partial V}{\partial h}\right) = \frac{\partial}{\partial r}(\pi r^2) = 2\pi r$$

$$\frac{\partial^2 V}{\partial h\,\partial r} = \frac{\partial}{\partial h}\left(\frac{\partial V}{\partial r}\right) = \frac{\partial}{\partial h}(2\pi rh) = 2\pi r$$

It will be noted that, in this simple case, $\dfrac{\partial^2 V}{\partial r\,\partial h} = \dfrac{\partial^2 V}{\partial h\,\partial r}$, a result that is true for all those functions of two independent variables which are likely to be encountered by the student

INTEGRAL CALCULUS

Integration

Integration is the reverse process to differentiation

If $$\frac{dy}{dx} = f'(x)$$

then y is the "integral" of $f'(x)$

This is written as

$$y = \int f'(x)\,dx$$

There is no complete set of rules for integration In fact, in certain cases the results are unknown The process depends upon remembering the results of differentiation

Consider the integral of x^n

From a knowledge of differentiation

$$\frac{d}{dx}(x^{n+1}) = (n+1)\,x^n$$

Hence $$\frac{d}{dx}\left(\frac{1}{n+1}\,x^{n+1}\right) = \frac{n+1}{n+1}\,x^n = x^n$$

Thus $$\int x^n\,dx = \frac{1}{n+1}\,x^{n+1} + C \qquad (117)$$

The constant C occurs in the complete integral; for, when differentiated, any constant becomes zero.

Equation 117 is true in all cases except when $n = -1$.

It will be remembered that:

$$\frac{d}{dx}(\log_e x) = \frac{1}{x}$$

Hence
$$\int x^{-1}\,dx = \log_e x + C \tag{118}$$

Example.—

Integrate $5x^3 + 3$.

The x^3 must have come from an x^4 term; x^4 itself when differentiated produces $4x^3$ and not $5x^3$. Hence the integral of the first term must be $\frac{5}{4}x^4$. The integral of the second term is similarly $3x$. The complete integral is $\frac{5}{4}x^4 + 3x + C$. *Ans.*

Consider the integral of $(ax + b)^n$.

$$\frac{d}{dx}(ax + b)^{n+1} = a(n + 1)(ax + b)^n$$

Hence
$$\frac{d}{dx}\left\{\frac{1}{a(n+1)}(ax + b)^{n+1}\right\} = (ax + b)^n$$

Thus
$$\int (ax + b)^n\,dx = \frac{1}{a(n+1)}(ax + b)^{n+1} + C \tag{119}$$

Example.—

Integrate $(2x + 3)^4$.

The answer may be written down straight away from equation 119.

$$\int (2x + 3)^4\,dx = \frac{1}{2 \times 5}(2x + 3)^5 + C$$
$$= \frac{1}{10}\cdot(2x + 3)^5 + C \quad \textit{Ans.}$$

The general rule for integration is to manipulate the integral into such a form that it is directly integrable using equations 117, 118 or 119.

Expansion

Frequently an expression can be converted into an integrable form by expanding.

Consider the integral of $(x^2 + 3)(x - 5)$.

$$\int (x^2 + 3)(x - 5)\,dx = \int (x^3 - 5x^2 + 3x - 15)\,dx$$
$$= \frac{x^4}{4} - \frac{5}{3}x^3 + \frac{3}{2}x^2 - 15x + C.$$

Integrals resulting in logs

If an expression $\log_e f(x)$ is differentiated, the result is $\dfrac{f'(x)}{f(x)}$ and any expression of this form may be integrated directly as

$\log_e f(x)$ That is, *if the numerator of an expression is the differential coefficient of the denominator, then the integral is the logarithm (to base e) of the denominator*

Consider the integral of $\tan x$

$$\int \tan x\, dx = \int \frac{\sin x}{\cos x}\, dx = -\int \frac{(-\sin x)}{\cos x}\, dx = -\log_e \cos x + C$$

Example

Integrate $\dfrac{12x+4}{3x^2+2x+6}$

In this case the numerator is twice the differential coefficient of the denominator

$$\int \frac{12x+4}{3x^2+2x+6}\, dx = 2\int \frac{6x+2}{3x^2+2x+6}\, dx$$

$$= 2\log_e (3x^2+2x+6) + C \qquad \textit{Ans}$$

Partial fractions

When the denominator of an expression may be factorised, the integral may often be found by rearranging the expression as partial fractions

Consider the integral of $\dfrac{x+1}{x^2+5x+6}$

$$\frac{x+1}{x^2+5x+6} = \frac{x+1}{(x+3)(x+2)} = \frac{2}{x+3} - \frac{1}{x+2}$$

Thus
$$\int \frac{x+1}{x^2+5x+6}\, dx = \int\left[\frac{2}{x+3} - \frac{1}{x+2}\right] dx$$

$$= \int \frac{2}{x+3}\, dx - \int \frac{1}{x+2}\, dx$$

$$= 2\log_e (x+3) - \log_e (x+2) + C$$

$$= \log_e \frac{(x+3)^2}{(x+2)} + C$$

Integration by parts

This is the converse of the rule for differentiating a product, and is used when the expression consists of the product of two different types of function of x, *e.g.* $x^3 \log x$

It will be remembered that

$$\frac{d}{dx}(uv) = u\frac{dv}{dx} + v\frac{du}{dx} \quad \text{(where } u \text{ and } v \text{ are functions of } x\text{)}$$

$$uv = \int\left(u\frac{dv}{dx}\right) dx + \int\left(v\frac{du}{dx}\right) dx$$

Let
$$w = \frac{dv}{dx} \qquad \therefore v = \int w\,.\,dx$$

Hence $u\int w\,dx = \int u\,w\,dx + \int\left(\int w\,dx\,\frac{du}{dx}\right)dx$

$$\therefore \int u\,w\,dx = u\int w\,dx - \int\left(\int w\,dx\,\frac{du}{dx}\right)dx \qquad (120)$$

Expressed in words, equation 120 becomes *the integral of a product equals the first term times the integral of the second term, minus the integral of (the integral of the second term times the differential of the first)*

Consider the integral of $x^3 \log_e x$

Since the differential coefficient of $\log_e x$ is known but the integral of $\log_e x$ has not yet been encountered take $w = x^3$ and $u = \log_e x$

Hence $\int x^3 \log_e x\,dx = \int \log_e x \cdot x^3\,dx$

$$= \log_e x \int x^3\,dx - \int\left(\int x^3\,dx \cdot \frac{d}{dx}\log_e x\right)dx$$

$$= \log_e x \cdot \frac{x^4}{4} - \int\left(\frac{x^4}{4}\cdot\frac{1}{x}\right)dx$$

$$= \frac{x^4}{4}\log_e x - \int\frac{x^3}{4}\,dx$$

$$= \frac{x^4}{4}\log_e x - \frac{x^4}{16}$$

For the complete solution a constant C must be added

Hence $\int x^3 \log_e x\,dx = \frac{x^4}{4}\log_e x - \frac{x^4}{16} + C$

This method may be used to determine the integral of $\log_e x$ by letting –

$$w = 1 \text{ and } u = \log_e x$$

Hence $\int \log_e x\,dx = \log_e x \int 1\,dx - \int\left(\int 1\,dx \cdot \frac{d}{dx}\log_e x\right)dx$

$$= x\log_e x - \int x \cdot \frac{1}{x}\,dx$$

$$= x\log_e x - x$$

Adding a constant for the complete solution gives –

$$\int \log_e x\,dx = x\log_e x - x + C$$

Trigonometrical transformations

Products and powers of sin and cos may be integrated by changing to the form $\sin nx$ and $\cos nx$

Consider the integral of $\sin^2 x$.

This cannot be integrated in its present form, hence apply the identity —

$$\cos 2x = 1 - 2\sin^2 x$$

Thus $\quad \sin^2 x = \frac{1}{2}(1 - \cos 2x)$

Hence $\quad \int \sin^2 x\, dx = \int \frac{1}{2}(1 - \cos 2x)\, dx = \frac{1}{2}\int(1 - \cos 2x)\, dx$

$$= \frac{1}{2}\int 1\, dx - \frac{1}{2}\int \cos 2x\, dx$$

$$= \frac{x}{2} - \frac{\sin 2x}{4} + C$$

Integration by substitution

The process of integration may frequently be simplified by making a substitution. Such a substitution may be either algebraic or trigonometric. It is important to remember that the "dx" term must be converted into terms of the new variable.

Algebraic substitutions. Consider the integral of $x(4 - 2x^2)^6$.

This may be determined by making an algebraic substitution.

Let $\quad 4 - 2x^2 = u$

Then $\quad -4x\, dx = du$

Hence $\quad \int x(4 - 2x^2)^6\, dx = \int -\frac{1}{4} u^6\, du$

$$= -\frac{u^7}{28} + C$$

$$= -\frac{(4 - 2x^2)^7}{28} + C$$

Trigonometrical substitutions.—If the expression to be integrated contains

(i) $\sqrt{a^2 - x^2}$ put $x = a \sin \theta$ or $x = a \cos \theta$

(ii) $\sqrt{a^2 + x^2}$ put $x = a \tan \theta$

(iii) $\sqrt{x^2 - a^2}$, put $x = a \sec \theta$

Consider the integral of $\dfrac{1}{\sqrt{a^2 - x^2}}$

This root can conveniently be removed by putting $x = a \sin \theta$, so that —

$$\sqrt{a^2 - x^2} = \sqrt{a^2 - a^2 \sin^2 \theta} = \sqrt{a^2 \cos^2 \theta} = a \cos \theta.$$

The "dx" has to be turned into an expression involving θ.

As $\quad x = a \sin \theta$

$$\therefore \quad \frac{dx}{d\theta} = a \cos \theta$$

$$dx = a \cos \theta\, d\theta$$

(This line can be obtained without the intermediate step.)

Hence $\quad \int \dfrac{dx}{\sqrt{a^2 - x^2}} = \int \dfrac{a \cos \theta\, d\theta}{a \cos \theta} = \int 1 \,.\, d\theta = \theta + C$

But $\quad \sin\theta = \dfrac{x}{a} \qquad \therefore\ \theta = \sin^{-1}\dfrac{x}{a}$

Thus $\quad \displaystyle\int \frac{dx}{\sqrt{a^2 - x^2}} = \sin^{-1}\frac{x}{a} + C$

Example.—

Find the integral of $\dfrac{\sqrt{16 - x^2}}{x^2}$

Put $x = 4\cos\theta, \quad \therefore\ \theta = \cos^{-1}\dfrac{x}{4}$

$$\therefore\ dx = -4\sin\theta\, d\theta.$$

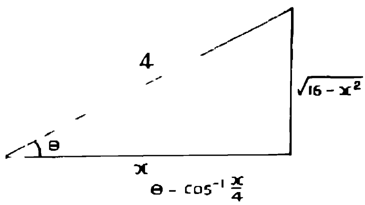


FIG. 62.

In addition, from Fig. 62

$$\sin\theta = \frac{\sqrt{16 - x^2}}{4}, \quad \cos\theta = \frac{x}{4}, \quad \tan\theta = \frac{\sqrt{16 - x^2}}{x}$$

$$\text{Hence} \int \frac{\sqrt{16 - x^2}}{x^2}\,dx = \int \frac{\sqrt{16 - 16\cos^2\theta}}{16\cos^2\theta}\,(-4\sin\theta\, d\theta)$$

$$= -\int \frac{\sin^2\theta}{\cos^2\theta}\,d\theta$$

$$= -\int \frac{1 - \cos^2\theta}{\cos^2\theta}\,d\theta$$

$$= -\int \sec^2\theta\, d\theta + \int 1\, d\theta$$

$$= -\tan\theta + \theta + C$$

$$= -\frac{\sqrt{16 - x^2}}{x} + \cos^{-1}\frac{x}{4} + C \quad \textit{Ans.}$$

Area under a curve

The area $OMPN$ under the curve $y = f(x)$ and bounded by the axes and the ordinate NP is a function of x (the co-ordinate of P) (*see* Fig. 63). The area cannot be found at once, but its derivatives can. For if x is increased to $(x + \delta x)$, the area A increases by an element δA, which is equal to the area $PP'N'N$. To a first approximation, this is equal to $y \cdot \delta x$.

$$\therefore\ \delta A = y\,\delta x \quad \text{or} \quad \frac{\delta A}{\delta x} = y$$

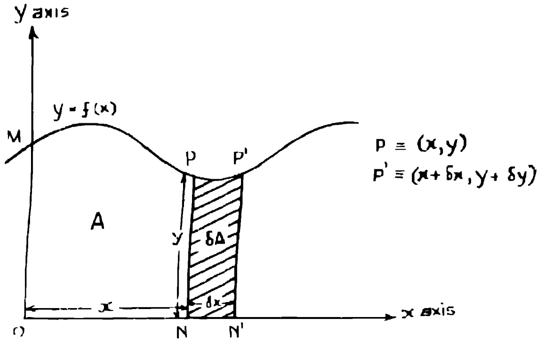


FIG. 63.—Infinitely small portion of the area under a curve.

$\therefore$ taking the limit as $\delta x \to 0$, the inaccuracy disappears and :—

$$\frac{dA}{dx} = y = f(x)$$

or
$$A = \int f(x)\,dx \qquad (121)$$

Hence the area of a curve up to the point x is found by integrating.

Definite integrals

Normally, the area is required between two ordinates, as at x_1 and x_2 in Fig. 64. This is found by calculating the area up to x_2 and subtracting from it the area up to x_1. If $A = \int f(x)\,dx$,

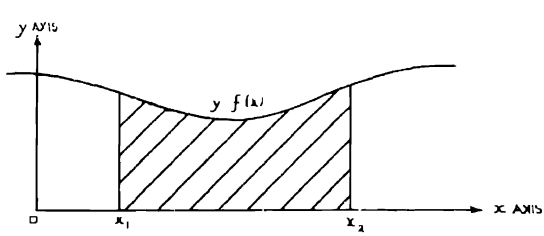


FIG. 64.—Finite area under a curve.

the notation used for this is $[A]_{x_1}^{x_2}$ *i.e.*, the value of A when $x = x_2$ minus the value of A when $x = x_1$. This is also written $\int_{x_1}^{x_2} f(x)\,dx$, and is known as a definite integral.

Note that as a result of subtraction the constant of integration disappears.

Example. –

Find the area under the curve $y = 3x^2 + 8x + 7$ from $x = 1$ to $x = 2$,

$$i.e., \int_1^2 (3x^2 + 8x + 7)\,dx$$

Evaluating the definite integral :—

Required area $= [x^3 + 4x^2 + 7x]_1^2$

$= (8 + 16 + 14) - (1 + 4 + 7) = 38 - 12 = 26$

Substitutions in definite integrals.—Note that if the variable is changed in order to simplify the integration, the answer need not be turned back into terms of the original variable, provided that the limits are changed to correspond with the substitution.

Example.—

Find the area of the circle $x^2 + y^2 = r^2$. It is easiest to calculate the area of one quadrant, taking $y = \sqrt{r^2 - x^2}$ from $x = 0$ to $x = r$, *i.e.*, $\int_0^r \sqrt{r^2 - x^2}\, dx$

A substitution has to be made :—

let $\qquad x = r \sin \theta$

$\therefore \qquad dx = r \cos \theta \, d\theta$

and $\qquad \sqrt{r^2 - x^2} = \sqrt{r^2 - r^2 \sin^2 \theta} = r \cos \theta$

Therefore the integral is $\int_{x=0}^{x=r} r \cos \theta \, . \, r \cos \theta \, d\theta = r^2 \int_{x=0}^{x=r} \cos^2 \theta \, d\theta$

The limits are in terms of x, and must be changed to θ.

$x = r$ corresponds to $\theta = \frac{\pi}{2}$; $x = 0$ to $\theta = 0$.

$$\therefore \qquad \text{area} = r^2 \int_0^{\frac{\pi}{2}} \cos^2 \theta \, d\theta$$

To integrate $\cos^2 \theta$, use the identity $\cos^2 \theta = \frac{1}{2}(1 + \cos 2\theta)$

$$\therefore \int = \tfrac{1}{2} r^2 \int_0^{\frac{\pi}{2}} (1 + \cos 2\theta)\, d\theta$$

$$= \frac{r^2}{2} \left[\theta + \frac{\sin 2\theta}{2} \right]_0^{\frac{\pi}{2}}$$

$$= \frac{r^2}{2} \left[\frac{\pi}{2} + 0 - 0 - 0 \right]$$

$$= \frac{\pi r^2}{4}$$

This is the area of $\frac{1}{4}$ of the circle. Hence the total area $= \pi r^2$.

Mean height of curve.—The mean height of a curve is obtained by finding the area of the curve and dividing it by the length of the base.

Example.—

Find the mean height of the curve $y = A^2 \sin^2 \theta$. If this curve be drawn, it is seen to extend along the θ-axis from $-\infty$ to $+\infty$. It is, however, *recurrent*, so that it is sufficient to consider the portion of the curve between $\theta = 0$, and $\theta = 2\pi$; outside these limits, the curve merely repeats the cycle of values it assumes within them. Its mean height can therefore be found by finding the area over one complete cycle (*e.g.*, from $\theta = 0$ to $\theta = 2\pi$, or equally, from $-\pi$ to $+\pi$, or from -2π to 0) and dividing by the length of the base (2π).

$$y = A^2 \sin^2 \theta$$

$\therefore$ the area over one cycle is given by

$$\int_0^{2\pi} y\, d\theta = \int_0^{2\pi} A^2 \sin^2 \theta\, d\theta$$

$$= \tfrac{1}{2}A^2 \int_0^{2\pi} (1 - \cos 2\theta)\, d\theta$$

$$= \tfrac{1}{2}A^2 \Big[\theta\Big]_0^{2\pi} - \tfrac{1}{2}A^2 \Big[\tfrac{1}{2}\sin 2\theta\Big]_0^{2\pi}$$

$$= \tfrac{1}{2}A^2 [2\pi - 0] - \tfrac{1}{2}A^2 [0 - 0]$$

$$= \pi A^2$$

The mean height h of the curve is therefore :—

$$h = \frac{\pi A^2}{2\pi}$$

$$= \frac{A^2}{2}$$

This might also be expressed by saying that $\frac{A^2}{2}$ is the mean or average value of the quantity $A^2 \sin^2 \theta$.

CIRCULAR AND HYPERBOLIC FUNCTIONS

Maclaurin's theorem

Many functions of x can be expanded as a series of powers of x (*e.g.*, $(1 + x)^n$, e^x, *etc.*). Maclaurin's theorem enables an expansion to be found for a general function of x, *i.e.*, $y = f(x)$.

Let $\quad f(x) = A_0 + A_1x + A_2x^2 + A_3x^3 + A_4x^4 + \ldots$

Differentiating :—

$$f'(x) = A_1 + 2 \,.\, A_2x + 3 \,.\, A_3x^2 + 4 \,.\, A_4x + \ldots$$
$$f''(x) = 2 \,.\, A_2 + 2 \,.\, 3 \,.\, A_3x + 3 \,.\, 4 \,.\, A_4x^2 + \ldots$$
$$f'''(x) = 2 \,.\, 3 \,.\, A_3 + 2 \,.\, 3 \,.\, 4 \,.\, A_4x + \ldots$$
$$f''''(x) = 2 \,.\, 3 \,.\, 4 \,.\, A_4 + \ldots$$

This is true for all values of x, hence, when $x = 0$.—

$$A_0 = f(0)$$

$$A_1 = f'(0)$$

$$A_2 = \frac{f''(0)}{2} = \frac{f''(0)}{\underline{|2}}$$

$$A_3 = \frac{f'''(0)}{2.3} = \frac{f'''(0)}{\underline{|3}}$$

$$A_4 = \frac{f''''(0)}{2.3.4} = \frac{f''''(0)}{\underline{|4}}$$

Thus

$$f(x) = f(0) + xf'(0) + \frac{x^2}{\underline{|2}} f''(0) + \frac{x^3}{\underline{|3}} f'''(0) + \frac{x^4}{\underline{|4}} f''''(0) + \ldots \quad (122)$$

where $f''(0)$ *etc.*, means the value of $f''(x)$, *etc.*, when $x = 0$. This is known as "*Maclaurin's theorem*."

If no derivatives of $f(x)$ vanish this expansion will involve an infinite number of terms. The reader should verify that it holds for $(1 + x)^n$ and e^x.

Circular functions

A number of important series can be obtained from this theorem.

Take for example, $f(x) = \sin x$. To find the series, one must calculate the successive derivatives and their value when $x = 0$.

$f(x)$	$= \sin x$	$f(0)$	$= 0$
$f'(x)$	$= \cos x$	$f'(0)$	$= 1$
$f''(x)$	$= -\sin x$	$f''(0)$	$= 0$
$f'''(x)$	$= -\cos x$	$f'''(0)$	$= -1$
$f''''(x)$	$= \sin x$	$f''''(0)$	$= 0$, *etc.*

Hence the series is.—

$$\sin x = x - \frac{x^3}{\underline{|3}} + \frac{x^5}{\underline{|5}} - \frac{x^7}{\underline{|7}} + \ldots \quad (123)$$

Similarly

$$\cos x = 1 - \frac{x^2}{\underline{|2}} + \frac{x^4}{\underline{|4}} - \frac{x^6}{\underline{|6}} + \ldots \quad (124)$$

Since $\quad e^x = 1 + x + \frac{x^2}{\underline{|2}} + \frac{x^3}{\underline{|3}} + \ldots$

the expansion for e^{jx} is —

$$e^{jx} = 1 + jx + \frac{(jx)^2}{\underline{|2}} + \frac{(jx)^3}{\underline{|3}} + \ldots$$

$$\therefore \quad e^{jx} = 1 + jx - \frac{x^2}{\underline{|2}} - j\frac{x^3}{\underline{|3}} + \frac{x^4}{\underline{|4}} + j\frac{x^5}{\underline{|5}} - \frac{x^6}{\underline{|6}} \quad \ldots$$

$$\therefore \quad e^{jx} = \left(1 - \frac{x^2}{\underline{|2}} + \frac{x^4}{\underline{|4}} \ldots\right) + j\left(x - \frac{x^3}{\underline{|3}} + \frac{x^5}{\underline{|5}} \ldots\right)$$

Hence $\qquad e^{jx} = \cos x + j \sin x \qquad (125)$

Similarly

$$e^{-jx} = \cos x - j \sin x \qquad (126)$$

This shows that the trigonometrical ratios can be treated from an algebraical aspect as well as from a geometrical aspect. It is possible to prove trigonometrical identities from these results, *e.g.*, multiplying together the two equations just obtained gives:—

$$1 = \cos^2 x + \sin^2 x$$

Both $\cos x$ and $\sin x$ may be obtained as expressions involving e by adding and subtracting these two equations. From $\cos x$ and $\sin x$, expressions for $\sec x$, $\operatorname{cosec} x$, $\tan x$, and $\cot x$ follow. Thus:—

$$\cos x = \frac{e^{jx} + e^{-jx}}{2} \qquad (127)$$

$$\sin x = \frac{e^{jx} - e^{-jx}}{j2} \qquad (128)$$

Hence $\qquad \tan x = j \cdot \frac{(e^{jx} - e^{-jx})}{(e^{jx} + e^{-jx})} \qquad (129)$

$$\sec x = \frac{2}{e^{jx} + e^{-jx}} \qquad (130)$$

$$\operatorname{cosec} x = \frac{j2}{e^{jx} - e^{-jx}} \qquad (131)$$

$$\cot x = j\frac{e^{jx} + e^{-jx}}{e^{jx} - e^{-jx}} \qquad (132)$$

Hyperbolic functions

It has been shown that $\cos x = \frac{e^{jx} + e^{-jx}}{2}$. The value of $\frac{e^{x} + e^{-x}}{2}$ is also important, and this is known as "$\cosh x$", or the "hyperbolic cosine" of x.

Thus: $\qquad \cosh x = \frac{e^{x} + e^{-x}}{2} \qquad (133)$

Similarly: $\qquad \sinh x = \frac{e^{x} - e^{-x}}{2} \qquad (134)$

Whence: $\quad \cosh x + \sinh x = e^{x} \qquad (135)$

and $\qquad \cosh x - \sinh x = e^{-x} \qquad (136)$

These hyperbolic functions bear the same relation to a hyperbola as sine and cosine do to a circle (*see* Appendix I).

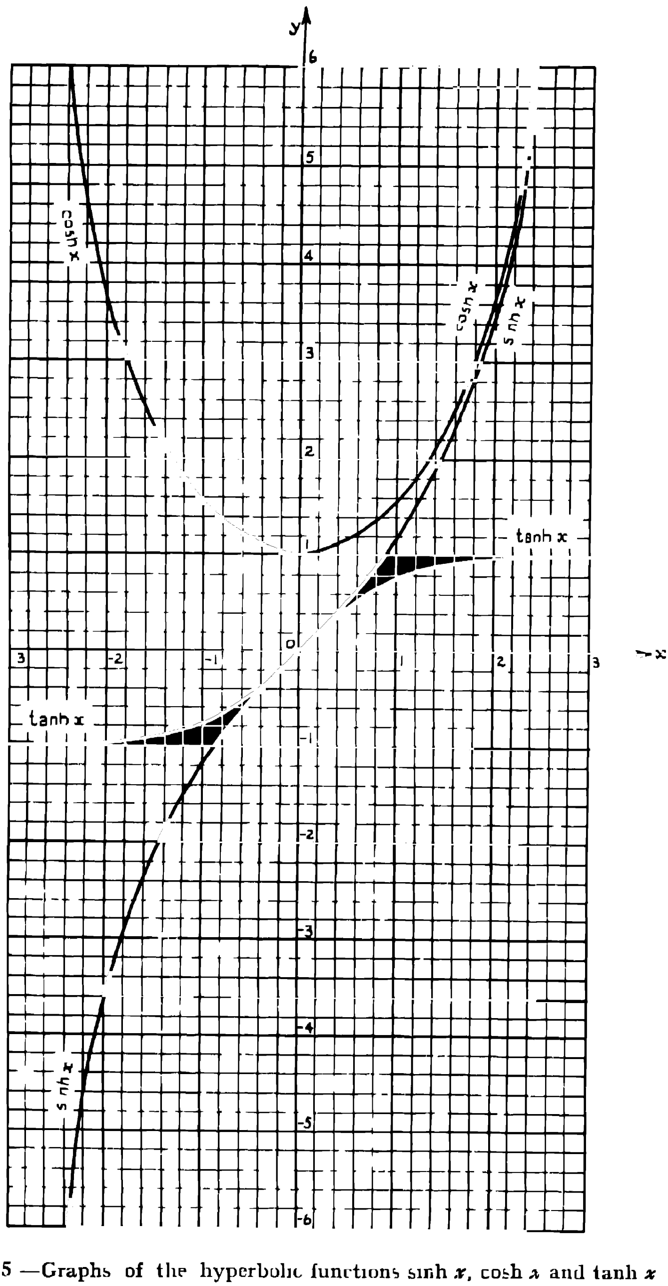


FIG 65 —Graphs of the hyperbolic functions sinh x, cosh x and tanh x

Since :— $$e^x = 1 + x + \frac{x^2}{\underline{|2}} + \frac{x^3}{\underline{|3}} + \frac{x^4}{\underline{|4}} + \ldots$$

and $$e^{-x} = 1 - x + \frac{x^2}{\underline{|2}} - \frac{x^3}{\underline{|3}} + \frac{x^4}{\underline{|4}} - \ldots$$

it follows that the series for the hyperbolic functions are :—

$$\cosh x = 1 + \frac{x^2}{\underline{|2}} + \frac{x^4}{\underline{|4}} + \ldots \qquad (137)$$

$$\sinh x = x + \frac{x^3}{\underline{|3}} + \frac{x^5}{\underline{|5}} + \ldots \qquad (138)$$

From the definitions, it can be seen that the following conversion rules apply :—

Circular to hyperbolic		*Hyperbolic to circular*	
$\sin x = -j \, . \sinh jx$	(139)	$\sinh x = -j \, . \sin jx$	(143)
$\cos x = \cosh jx$	(140)	$\cosh x = \cos jx$	(142)
$\sin jx = j \, . \sinh x$	(141)	$\sinh jx = j \, . \sin x$	(144)
$\cos jx = \cosh x$	(142)	$\cosh jx = \cos x$	(140)

Tanh x is defined as $$\frac{\sinh x}{\cosh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{e^{2x} - 1}{e^{2x} + 1} \qquad (145)$$

therefore $$\tanh x = \frac{j \sin jx}{\cos jx} = -j \tan jx \qquad (146)$$

and $$\tan x = -j \tanh jx \qquad (147)$$

Fig. 65 gives the graphs of the hyperbolic functions.

Note that cosh x is always greater than 1, and that tanh x lies between $+1$ and -1; and that, unlike circular functions, hyperbolic functions are not periodic.

coth x is defined as $$\frac{\cosh x}{\sinh x} = \frac{1}{\tanh x} = \frac{e^x + e^{-x}}{e^x - e^{-x}} = \frac{e^{2x} + 1}{e^{2x} - 1} \qquad (148)$$

sech x is defined as $$\frac{1}{\cosh x} = \frac{2}{e^x + e^{-x}} \qquad (149)$$

cosech x is defined as $$\frac{1}{\sinh x} = \frac{2}{e^x - e^{-x}} \qquad (150)$$

Hyperbolic identities

Hyperbolic identities are similar to the corresponding circular identities, and may be readily deduced from them. As a general rule, identities hold if " $-\sinh^2$ " is written instead of " $\sin^2$ ", and " $\cosh^2$ " instead of " $\cos^2$ ".

Thus $\cos^2 x + \sin^2 x = 1$

becomes $$\cosh^2 x - \sinh^2 x = 1 \qquad (151)$$

and $\cos 2x = \cos^2 x - \sin^2 x$

becomes
$$\cosh 2x = \cosh^2 x + \sinh^2 x \qquad (152)$$
$$= 2\cosh^2 x - 1 \qquad (153)$$
$$= 1 + 2\sinh^2 x \qquad (154)$$

It can be shown that—

$$\sinh (A + B) = \sinh A \cosh B + \cosh A \sinh B \tag{155}$$

$$\sinh (A - B) = \sinh A \cosh B - \cosh A \sinh B \tag{156}$$

$$\cosh (A + B) = \cosh A \cosh B + \sinh A \sinh B \tag{157}$$

$$\cosh (A - B) = \cosh A \cosh B - \sinh A \sinh B \tag{158}$$

Dividing equation 155 by equation 157 gives

$$\tanh (A + B) = \frac{\tanh A + \tanh B}{1 + \tanh A \tanh B} \tag{159}$$

Dividing equation 156 by equation 158 gives

$$\tanh (A - B) = \frac{\tanh A - \tanh B}{1 - \tanh A \tanh B} \tag{160}$$

Putting $A = B = x$ in equation 155 gives

$$\sinh 2x = 2 \sinh x \cosh x \tag{161}$$

Putting $A = B = x$ in equation 157 gives equation 152 above.

It will be noted that

$$\begin{aligned} \tanh x &= \frac{\sinh x}{\cosh x} \\ &= \frac{2 \sinh^2 x}{2 \sinh x \cosh x} \\ &= \frac{\cosh 2x - 1}{\sinh 2x} \end{aligned} \tag{162}$$

All the above identities may be verified using equations 133 and 134.

Differentiation of hyperbolic functions

The differential coefficients of hyperbolic functions can easily be obtained by using the exponential form of the functions. Thus for example

$$\begin{aligned} \frac{d}{dx}(\sinh x) &= \frac{d}{dx}\left(\frac{e^x - e^{-x}}{2}\right) \\ &= \frac{d}{dx}\left(\tfrac{1}{2}e^x - \tfrac{1}{2}e^{-x}\right) \\ &= \tfrac{1}{2}e^x + \tfrac{1}{2}e^{-x} \\ &= \frac{e^x + e^{-x}}{2} \\ &= \cosh x \end{aligned} \tag{163}$$

Some of the more important derivatives are given in Table III, while a further list is given in Appendix I.

TABLE III

Differential coefficients of some hyperbolic functions

y	$\dfrac{dy}{dx}$
$\sinh x$	$\cosh x$
$\cosh x$	$\sinh x$
$\tanh x$	$\operatorname{sech}^2 x$
$\coth x$	$-\operatorname{cosech}^2 x$
$\operatorname{sech} x$	$-\operatorname{sech} x \tanh x$
$\operatorname{cosech} x$	$-\operatorname{cosech} x \coth x$
$\sinh^{-1} x$	$\dfrac{1}{\sqrt{1+x^2}}$
$\cosh^{-1} x$	$\dfrac{1}{\sqrt{x^2-1}}$
$\tanh^{-1} x$	$\dfrac{1}{1-x^2}$

Complex hyperbolic functions

Expressions such as $\cosh(\alpha + j\beta)$ are often encountered in transmission theory. Their values can be calculated from first principles and tables.

If $\sinh(\alpha + j\beta) = A + jB$, one can find A and B in terms of α and β.

For
$$\begin{aligned} A + jB &= \sinh(\alpha + j\beta) \\ &= \sinh\alpha \cosh j\beta + \cosh\alpha \sinh j\beta \\ &= \sinh\alpha \cos\beta + j\cosh\alpha \sin\beta \end{aligned}$$

Equating real and imaginary parts —
$$A = \sinh\alpha \cos\beta$$
and
$$B = \cosh\alpha \sin\beta$$

Similarly, if $\cosh(\alpha + j\beta) = A + jB$ one can find A and B in terms of α and β.

$$\begin{aligned} A + jB &= \cosh(\alpha + j\beta) \\ &= \cosh\alpha \cosh j\beta + \sinh\alpha \sinh j\beta \\ &= \cosh\alpha \cos\beta + j\sinh\alpha \sin\beta \end{aligned}$$

Equating real and imaginary parts
$$A = \cosh\alpha \cos\beta$$
and
$$B = \sinh\alpha \sin\beta$$

Example 1.—Evaluate $\cosh\left(3 + j\frac{\pi}{4}\right)$.

$$\begin{aligned} \cosh\left(3 + j\tfrac{\pi}{4}\right) &= \cosh 3 \cosh j\tfrac{\pi}{4} + \sinh 3 \sinh j\tfrac{\pi}{4} \\ &= \cosh 3 \cos\tfrac{\pi}{4} + j\sinh 3 \sin\tfrac{\pi}{4} \\ &= 10{\cdot}07 \times 0{\cdot}7071 + j\,10{\cdot}02 \times 0{\cdot}7071 \\ &= 7{\cdot}12 + j\,7{\cdot}07 \end{aligned}$$

Example 2.—Show that $\tanh\left(\alpha + j\frac{\pi}{2}\right) = \coth\alpha$

$$\sinh\left(\alpha + j\frac{\pi}{2}\right) = \sinh\alpha\cosh j\frac{\pi}{2} + \cosh\alpha\sinh j\frac{\pi}{2}$$
$$= \sinh\alpha\cos\frac{\pi}{2} + j\cosh\alpha\sin\frac{\pi}{2}$$
$$= j\cosh\alpha$$

and

$$\cosh\left(\alpha + j\frac{\pi}{2}\right) = \cosh\alpha\cosh j\frac{\pi}{2} + \sinh\alpha\sinh j\frac{\pi}{2}$$
$$= \cosh\alpha\cos\frac{\pi}{2} + j\sinh\alpha\sin\frac{\pi}{2}$$
$$= j\sinh\alpha$$

$$\therefore\ \tanh\left(\alpha + j\frac{\pi}{2}\right) = \frac{\sinh\left(\alpha + j\frac{\pi}{2}\right)}{\cosh\left(\alpha + j\frac{\pi}{2}\right)} = \frac{j\cosh\alpha}{j\sinh\alpha} = \coth\alpha \quad \text{Q.E.D.}$$

The converse of this type of problem is less simple—*i.e.*, given $\tanh(\alpha + j\beta) = A + jB$, find α and β.

If $\qquad \tanh(\alpha + j\beta) = A + jB,$

then $\qquad \tanh(\alpha - j\beta) = A - jB,$

since if an identity is true for $+j$, it is also true for $-j$.

$$\tanh[(\alpha + j\beta) + (\alpha - j\beta)] = \frac{\tanh(\alpha + j\beta) + \tanh(\alpha - j\beta)}{1 + \tanh(\alpha + j\beta)\,.\,\tanh(\alpha - j\beta)}$$

$$\therefore \qquad \tanh 2\alpha = \frac{A + jB + A - jB}{1 + (A + jB)(A - jB)}$$
$$= \frac{2A}{1 + A^2 + B^2}$$

Also

$$\tanh[(\alpha + j\beta) - (\alpha - j\beta)] = \frac{\tanh(\alpha + j\beta) - \tanh(\alpha - j\beta)}{1 - \tanh(\alpha + j\beta)\,.\,\tanh(\alpha - j\beta)}$$

$$\therefore \qquad \tanh 2j\beta = \frac{A + jB - A + jB}{1 - (A + jB)(A - jB)}$$
$$= \frac{2jB}{1 - A^2 - B^2}$$

$$\therefore \qquad -j\tanh 2j\beta = \frac{2B}{1 - A^2 - B^2}$$

$$\therefore\ \tan 2\beta = \frac{2B}{1 - A^2 - B^2}$$

Thus
$$\tanh 2\alpha = \frac{2A}{1 + A^2 + B^2} \tag{164}$$

and
$$\tan 2\beta = \frac{2B}{1 - (A^2+B^2)} \tag{165}$$

Those roots for β must be chosen for which $\tan \beta$ has the same sign as B

If the hyperbolic function of a complex number is required in the polar form $r\angle\theta$ the following identities may be used –

$$\sinh(\alpha + j\beta) = \sqrt{\sinh^2\alpha + \sin^2\beta}\,\angle\underline{\tan^{-1}(\coth\alpha \tan\beta)} \tag{166}$$

$$\cosh(\alpha + j\beta) = \sqrt{\sinh^2\alpha + \cos^2\beta}\,\angle\underline{\tan^{-1}(\tanh\alpha \tan\beta)} \tag{167}$$

These identities may be verified as follows

Let $\sinh(\alpha + j\beta)$ equal a vector $r\angle\theta$

$$\begin{aligned}
\sinh(\alpha + j\beta) &= \sinh\alpha \cosh j\beta + \cosh\alpha \sinh j\beta \\
&= \sinh\alpha \cos\beta + j\cosh\alpha \sin\beta \\
r^2 &= \sinh^2\alpha \cos^2\beta + \cosh^2\alpha \sin^2\beta \\
&= \sinh^2\alpha (1 - \sin^2\beta) + (1 + \sinh^2\alpha)\sin^2\beta \\
&= \sinh^2\alpha - \sinh^2\alpha \sin^2\beta + \sin^2\beta + \sinh^2\alpha \sin^2\beta \\
&= \sinh^2\alpha + \sin^2\beta \\
r &= \sqrt{\sinh^2\alpha + \sin^2\beta}
\end{aligned}$$

Also
$$\begin{aligned}
\theta &= \tan^{-1}\left(\frac{\cosh\alpha}{\sinh\alpha} \cdot \frac{\sin\beta}{\cos\beta}\right) \\
&= \tan^{-1}(\coth\alpha \tan\beta)
\end{aligned}$$

Similarly letting $\cosh(\alpha + j\beta)$ equal $r\angle\theta$

$$\begin{aligned}
\cosh(\alpha + j\beta) &= \cosh\alpha \cosh j\beta + \sinh\alpha \sinh j\beta \\
&= \cosh\alpha \cos\beta + j\sinh\alpha \sin\beta \\
r^2 &= \cosh^2\alpha \cos^2\beta + \sinh^2\alpha \sin^2\beta \\
&= (1 + \sinh^2\alpha)\cos^2\beta + \sinh^2\alpha (1 - \cos^2\beta) \\
&= \cos^2\beta + \sinh^2\alpha \cos^2\beta + \sinh^2\alpha - \sinh^2\alpha \cos^2\beta \\
&= \sinh^2\alpha + \cos^2\beta \\
r &= \sqrt{\sinh^2\alpha + \cos^2\beta}
\end{aligned}$$

and
$$\begin{aligned}
\theta &= \tan^{-1}\left(\frac{\sinh\alpha}{\cosh\alpha} \cdot \frac{\sin\beta}{\cos\beta}\right) \\
&= \tan^{-1}(\tanh\alpha \tan\beta)
\end{aligned}$$

Alternatively the following identities may be used —

$$\sinh(\alpha + j\beta) = \sqrt{\cosh^2\alpha - \cos^2\beta}\,\angle\underline{\tan^{-1}(\coth\alpha \tan\beta)} \tag{168}$$

$$\cosh(\alpha + j\beta) = \sqrt{\cosh^2\alpha - \sin^2\beta}\,\angle\underline{\tan^{-1}(\tanh\alpha \tan\beta)} \tag{169}$$

DIFFERENTIAL EQUATIONS

A differential equation is an equation involving unknown quantities and their derivatives.

e.g. $$\frac{d^2y}{dx^2} + 5\frac{dy}{dx} + 4y = 0$$

and $$x^2\frac{dy}{dx} = \sin x.$$

A solution of the equation is a function whose derivatives satisfy the equation.

Equations containing first derivatives are known as "first-order" equations; equations containing first and second derivatives are known as "second-order" equations, and so on. In general, there will be one indeterminate constant in the solution of an equation containing only first derivatives, two for an equation containing second derivatives, and so on.

Solution of equations of the first order and first degree

If the first order equation $\left(\text{containing only } \frac{dy}{dx}\right)$ can be manipulated into the form $f(x)\,dx = F(y)\,dy$, it may be solved by direct integration

Example $$2y + 3\frac{dy}{dx} = 4$$

$$\therefore \quad 3\frac{dy}{dx} = 4 - 2y$$

$$\therefore \quad \frac{dy}{y-2} = -\frac{2}{3}dx$$

Integrating:

$$\log_e(y-2) = -\frac{2}{3}x + \log_e C \text{ (where } C \text{ is a constant)}$$

$$\therefore \quad y - 2 = Ce^{-\frac{2x}{3}}$$

or $$y = 2 + Ce^{-\frac{2x}{3}} \quad \textit{Ans.}$$

Linear equations of the first order, of the general form:—

$$\frac{dy}{dx} + Py = Q, \tag{170}$$

where P and Q may be functions of x, may be solved by multiplying through by e raised to the power of the integral, with respect to x, of the coefficient of y, *i.e.* by $e^{\int P dx}$.

$$\therefore \qquad e^{\int P dx} \cdot \frac{dy}{dx} + yPe^{\int P dx} = Qe^{\int P dx}$$

The left-hand side of the equation is the differential coefficient with respect to x of the product of $e^{\int P dx}$ and y.

$$\therefore \qquad \frac{d}{dx}\left(e^{\int P dx} \cdot y\right) = Q\, e^{\int P dx}$$

Integrating —

$$y\, e^{\int P dx} = \int \left[Q\, e^{\int P dx}\right] dx + C \qquad (171)$$

Example —

Solve $$\frac{dy}{dx} - 2y = e^{3x}$$

Multiply both sides by e^{-2x} which is e raised to the power of the integral with respect to x of (-2)

$$e^{-2x}\left(\frac{dy}{dx} - 2y\right) = e^{-2x} e^{3x}$$

$$\frac{d}{dx}\left(e^{-2x} y\right) = e^{x}$$

Integrating

$$e^{-2x} y = e^{x} + C$$

$$\therefore \qquad y = e^{2x}\left(e^{x} + C\right) \qquad Ans.$$

Solution of linear equations having constant coefficients

The equations considered here are linear equations of any order having constant coefficients, i.e. equations of the form —

$$p_n \frac{d^n y}{dx^n} + p_{n-1}\frac{d^{n-1} y}{dx^{n-1}} + \dots + p_1 \frac{dy}{dx} + p_0 y = f(x) \qquad (172)$$

where p_n etc. are constants.

The simple case where $f(x) = 0$ will be considered first, e.g. —

$$\frac{d^3 y}{dx^3} + 7\frac{d^2 y}{dx^2} + 9\frac{dy}{dx} + y = 0$$

The general form of such a linear equation is

$$p_n \frac{d^n y}{dx^n} + p_{n-1}\frac{d^{n-1} y}{dx^{n-1}} + \dots + p_1 \frac{dy}{dx} + p_0 y = 0 \qquad (173)$$

It can be shown that the general solution to this equation is —

$$y = A_1 e^{m_1 x} + A_2 e^{m_2 x} + \dots + A_n e^{m_n x} \qquad (174)$$

where $m_1, m_2, \dots, m_n$ are the solutions of the equation —

$$p_n m^n + p_{n-1} m^{n-1} + \dots + p_1 m + p_0 = 0$$

Note that this general solution contains n arbitrary constants.

***Example* 1**

Solve $$\frac{d^3 y}{dx^3} - 4\frac{dy}{dx} = 0$$

m_1, m_2, and m_3 will be the roots of :—

$$m^3 - 4m = 0$$

i.e. of
$$m(m-2)(m+2) = 0$$

$\therefore$
$$m = 0,\ 2, \text{ or } -2.$$

Therefore the general solution is :—

$$\begin{aligned} y &= Ae^0 + Be^{2x} + Ce^{-2x} \\ &= A + Be^{2x} + Ce^{-2x} \end{aligned}$$

Since $e^{2x} = \cosh 2x + \sinh 2x$ and $e^{-2x} = \cosh 2x - \sinh 2x$, this may be written as :—

$$\begin{aligned} y &= A + B(\cosh 2x + \sinh 2x) + C(\cosh 2x - \sinh 2x) \\ &= A + D \cosh 2x + E \sinh 2x \quad \textit{Ans.} \end{aligned}$$

where $D = B + C$, and $E = B - C$

Example 2.—

An important example in transmission theory is :—

$$\frac{d^2y}{dx^2} - \gamma^2 y = 0$$

The solution as above is :

$$\begin{aligned} y &= Ae^{\gamma x} + Be^{-\gamma x} \\ &= C \cosh \gamma x + D \sinh \gamma x \quad \textit{Ans.} \end{aligned}$$

where $C = A + B$, and $D = A - B$

Example 3.—

In some cases the roots may be imaginary, as in this example :—

$$\frac{d^2y}{dx^2} + 9y = 0$$

m_1 and m_2 are the roots of :—

$$m^2 + 9 = 0$$

$\therefore$
$$m = \pm j3$$

Therefore the general solution is :—

$$\begin{aligned} y &= Ae^{j3x} + Be^{-j3x} \\ &= A(\cos 3x + j \sin 3x) + B(\cos 3x - j \sin 3x) \\ &= C \cos 3x + jD \sin 3x \quad \textit{Ans} \end{aligned}$$

where $C = A + B$, and $D = A - B$

Consider now an equation of the form :

$$\frac{d^2y}{dx^2} + P\frac{dy}{dx} + Qy = f(x)$$

where P and Q are constants.

The complete solution to the equation may be written as :—

$$y = u + v$$

where u is any function whatever that satisfies the equation, and v is the general solution to the equation :—

$$\frac{d^2y}{dx^2} + P\frac{dy}{dx} + Qy = 0$$

u is called the "particular integral", and
v is called the "complementary function".

Example.—

Solve : $\dfrac{d^2y}{dx^2} + \dfrac{dy}{dx} - 6y = x^2$

To find the complementary function.—Consider the equation :—

$$\frac{d^2y}{dx^2} + \frac{dy}{dx} - 6y = 0$$

$$m^2 + m - 6 = 0$$

$$(m + 3)(m - 2) = 0$$

Hence $\qquad y = Ae^{-3x} + Be^{2x}$

This is the complementary function.

To find the particular integral.—

Let $y = H + Kx + Lx^2$ be one solution to the equation.

Then $\qquad \dfrac{dy}{dx} = K + 2Lx$

and $\qquad \dfrac{d^2y}{dx^2} = 2L$

These derivatives must satisfy the original equation ;

$$\therefore \quad 2L + (K + 2Lx) - 6(H + Kx + Lx^2) = x^2$$

$$\therefore \quad 2L + K + 2Lx - 6H - 6Kx - 6Lx^2 = x^2$$

Thus equating coefficients gives :—

$$L = -\frac{1}{6}; \quad K = -\frac{1}{18}; \quad H = -\frac{7}{108};$$

Hence the particular integral $= -\dfrac{1}{108}(7 + 6x + 18x^2)$

The complete solution is therefore :—

$$y = A\,.\,e^{-3x} + B\,.\,e^{2x} - \frac{1}{108}(7 + 6x + 18x^2)$$

HARMONIC ANALYSIS

Fourier's Theorem

Any single-valued periodic function $y = f(t)$ having a period 2π may be expressed in the form :—

$$y = c + A_1 \sin(\omega t + \varphi_1) + A_2 \sin(2\omega t + \varphi_2) + A_3 \sin(3\omega t + \varphi_3) + \ldots \quad (175)$$

Since $A \sin(\omega t + \varphi) = A \sin \omega t \cos \varphi + A \cos \omega t \sin \varphi$
$= a \sin \omega t + b \cos \omega t$

where $a = A \cos \varphi$ and $b = A \sin \varphi$,

this expansion may be expressed as :—

$$y = c + a_1 \sin \omega t + a_2 \sin 2\omega t + a_3 \sin 3\omega t + \ldots + b_1 \cos \omega t + b_2 \cos 2\omega t + b_3 \cos 3\omega t + \ldots \quad (176)$$

In order to use this expression to analyse a complex waveform, it is necessary to determine the coefficients c, a_1, a_2, . . . and b_1, b_2, ... This is done by multiplying both sides of equation 176 by a suitable factor and integrating between the limits 0 and 2π. If the multiplying factor is correctly chosen, all terms vanish except those which will give the required coefficient. The vanishing of the unknown terms depends on certain definite integrals:—

Integrals of harmonic functions.

(*a*) $$\int_0^{2\pi} \cos nx\, dx = \frac{1}{n}\Big[\sin nx\Big]_0^{2\pi} = \frac{1}{n}\Big[0 - 0\Big]$$
$$= 0 \text{ (n is any integer)} \qquad (177)$$

(*b*) $$\int_0^{2\pi} \sin nx\, dx = -\frac{1}{n}\Big[\cos nx\Big]_0^{2\pi} = -\frac{1}{n}\Big[1 - 1\Big]$$
$$= 0 \text{ (n is any integer)} \qquad (178)$$

(*c*) $$\int_0^{2\pi} \sin mx \cos nx\, dx = \int_0^{2\pi} \{\tfrac{1}{2}\sin(m+n)x + \tfrac{1}{2}\sin(m-n)x\}\, dx$$
$$= 0 \text{ (m and n are any integers)} \qquad (179)$$

(*d*) $$\int_0^{2\pi} \cos mx \cos nx\, dx = \int_0^{2\pi} \{\tfrac{1}{2}\cos(m+n)x + \tfrac{1}{2}\cos(m-n)x\}\, dx$$
$$= 0 \text{ (m and n are any unequal integers)} \qquad (180)$$

(*e*) $$\int_0^{2\pi} \sin mx \sin nx\, dx = \int_0^{2\pi} \{\tfrac{1}{2}\cos(m-n)x - \tfrac{1}{2}\cos(m+n)x\}\, dx$$
$$= 0 \text{ (m and n are any unequal integers)} \qquad (181)$$

(*f*) $$\int_0^{2\pi} \cos^2 nx\, dx = \int_0^{2\pi} \{\tfrac{1}{2}\cos 2nx + \tfrac{1}{2}\}\, dx$$
$$= \pi \text{ (n is any integer)} \qquad (182)$$

(*g*) $$\int_0^{2\pi} \sin^2 nx\, dx = \int_0^{2\pi} \{\tfrac{1}{2} - \tfrac{1}{2}\cos 2nx\}\, dx$$
$$= \pi \text{ (n is any integer)} \qquad (183)$$

Determination of the coefficients

To find c.—Integrate both sides of equation 176 with respect to ωt between the limits 0 and 2π.

$$\int_0^{2\pi} y\, d(\omega t) = c\int_0^{2\pi} d(\omega t) + a_1\int_0^{2\pi} \sin\omega t.d(\omega t)$$
$$+ a_2\int_0^{2\pi} \sin 2\omega t.d(\omega t) + \ldots$$
$$+ b_1\int_0^{2\pi} \cos\omega t.d(\omega t)$$
$$+ b_2\int_0^{2\pi} \cos 2\omega t.d(\omega t) + \ldots$$

$$= c\int_0^{2\pi} d(\omega t) \quad + 0$$
$$+ 0 + \ldots$$
$$+ 0$$
$$+ 0 + \ldots$$

$$\therefore \int_0^{2\pi} y\, d(\omega t) = c\int_0^{2\pi} d(\omega t)$$
$$= c\Big[\omega t\Big]_0^{2\pi}$$
$$= 2\pi\, c$$

$$\therefore \qquad c = \frac{1}{2\pi}\int_0^{2\pi} y\, d(\omega t) \tag{184}$$

It will be noted that c is the mean value of y between the limits 0 and 2π.

To find a_n, the series must be multiplied by some factor such that on integrating between the limits 0 and 2π, all terms vanish except that containing a_n. This may be accomplished by multiplying the series by $\sin n\omega t$ and integrating with respect to ωt between limits 0 and 2π.

$$\int_0^{2\pi} y \sin n\omega t \,.\, d(\omega t) = c\int_0^{2\pi} \sin n\omega t \,.\, d(\omega t)$$
$$+ a_1\int_0^{2\pi} \sin \omega t \sin n\omega t \,.\, d(\omega t)$$
$$+ a_2\int_0^{2\pi} \sin 2\omega t \sin n\omega t \,.\, d(\omega t) + \ldots$$
$$+ a_n\int_0^{2\pi} \sin^2 n\omega t \,.\, d(\omega t) + \ldots$$
$$+ b_1\int_0^{2\pi} \cos \omega t \sin n\omega t \,.\, d(\omega t)$$
$$+ b_2\int_0^{2\pi} \cos 2\omega t \sin n\omega t \,.\, d(\omega t) + \ldots$$
$$= 0$$
$$+ 0 + 0 + 0 \ldots + a_n.\pi + 0 \ldots$$
$$+ 0 + 0 + 0 \ldots$$
$$= a_n \,.\, \pi$$

Hence
$$a = \frac{1}{\pi}\int_0^{2\pi} y \sin n\omega t \,.\, d(\omega t) \tag{185}$$

where n is any positive integer.

In this case, it will be noticed that a_n is twice the mean value of $y \sin n\omega t$ between the limits of 0 and 2π.

To find b_n, the series is multiplied and integrated in a similar manner, the necessary factor being $\cos n\omega t$. The series then becomes :—

$$\begin{aligned}\int_0^{2\pi} y \cos n\omega t \,.\, d(\omega t) = c\int_0^{2\pi} & \cos n\omega t \,.\, d(\omega t)\\ &+ a_1\int_0^{2\pi} \sin \omega t \cos n\omega t \,.\, d(\omega t)\\ &+ a_2\int_0^{2\pi} \sin 2\omega t \cos n\omega t \,.\, d(\omega t) + \ldots\\ &+ b_1\int_0^{2\pi} \cos \omega t \cos n\omega t \,.\, d(\omega t)\\ &+ b_2\int_0^{2\pi} \cos 2\omega t \cos n\omega t \,.\, d(\omega t) + \ldots\\ &+ b_n\int_0^{2\pi} \cos^2 n\omega t \,.\, d(\omega t) + \ldots\\ = 0 &\\ &+ 0 + 0 + 0 \ldots\\ &+ 0 + 0 + 0 \ldots + b_n.\pi + 0 \ldots\\ = b_n.\pi &\end{aligned}$$

$$\therefore \quad b_n = \frac{1}{\pi}\int_0^{2\pi} y \cos n\omega t \,.\, d(\omega t) \tag{186}$$

where n is any positive integer.

Thus b_n is twice the mean value of $y \cos n\omega t$ between the limits 0 and 2π.

Analysis of a square waveform

Fig. 66 shows a square waveform ; this is a single-valued periodic function of ωt, having a period 2π, and it may therefore be analysed by Fourier's Theorem.

From $\omega t = 0$ to $\omega t = \pi$, the equation to the function is $y = d$
From $\omega t = \pi$ to $\omega t = 2\pi$, the equation to the function is $y = 0$

Let the function, expressed as a harmonic series, be :—

$$\begin{aligned} y = c &+ a_1 \sin \omega t + a_2 \sin 2\omega t + a_3 \sin 3\omega t + \ldots\\ &+ b_1 \cos \omega t + b_2 \cos 2\omega t + b_3 \cos 3\omega t + \ldots\end{aligned}$$

To find c.—

$$c = \frac{1}{2\pi}\int_0^{2\pi} y \, d(\omega t) = \frac{1}{2\pi}\int_0^{\pi} y \, d(\omega t) + \frac{1}{2\pi}\int_{\pi}^{2\pi} y \, d(\omega t)$$

$$\therefore \qquad c = \frac{1}{2\pi} \cdot \mathrm{d} \cdot \pi + 0 = \frac{\mathrm{d}}{2}.$$

From inspection of Fig. 66, it can be verified that the mean value of y is $\frac{\mathrm{d}}{2}$.

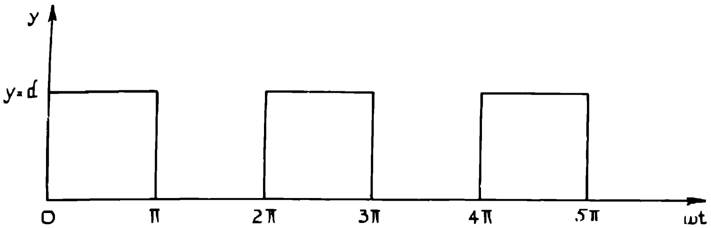


FIG. 66.—Square waveform always above t axis.

To find a_n.—

$$a_n = \frac{1}{\pi}\int_0^{2\pi} y \sin n\omega t \, . \, d\,(\omega t)$$

$$= \frac{1}{\pi}\left[\int_0^{\pi} \mathrm{d} \sin n\omega t \, . \, d(\omega t) + \int_{\pi}^{2\pi} 0 \, . \, \sin n\omega t \, . \, d(\omega t)\right]$$

$$= \frac{1}{\pi}\left[\frac{-\,\mathrm{d} \cos n\omega t}{n}\right]_0^{\pi}$$

$$= \frac{\mathrm{d}}{n\pi}(1 - \cos n\pi)$$

When n is odd, $(1 - \cos n\pi) = 2$

When n is even, $(1 - \cos n\pi) = 0$

Thus :—

$$a_1 = \frac{2\mathrm{d}}{\pi}, \quad a_2 = 0$$

$$a_3 = \frac{2\mathrm{d}}{3\pi}, \quad a_4 = 0$$

To find b_n.—

$$b_n = \frac{1}{\pi}\int_0^{2\pi} y \cos n\omega t \, . \, d(\omega t)$$

$$= \frac{1}{\pi}\left[\int_0^{\pi} \mathrm{d} \cos n\omega t \, . \, d(\omega t) + \int_{\pi}^{2\pi} 0 . \cos n\omega t \, . \, d(\omega t)\right]$$

$$= \frac{1}{\pi}\left[\frac{\mathrm{d} \sin n\omega t}{n}\right]_0^{\pi}$$

$$= 0$$

All cosine terms are thus zero.

The required equation to the function is therefore :—

$$y = \frac{d}{2} + \frac{2d}{\pi}\left\{\sin \omega t + \frac{1}{3}\sin 3\,\omega t + \frac{1}{5}\sin 5\,\omega t + \frac{1}{7}\sin 7\,\omega t + \ldots\right\} \quad (187)$$

A curve frequently encountered in communications engineering is that shown in Fig. 67. This is similar in general form to the one

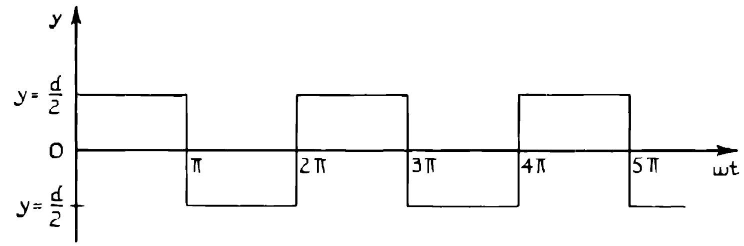


FIG. 67.—Square waveform with zero constant term.

shown in Fig. 66, and represented by equation 187, but it is symmetrical about the time axis (0–ωt); the first term $\left(\frac{d}{2}\right)$ therefore does not appear in its equation, which is :—

$$y = \frac{2d}{\pi}\left(\sin \omega t + \frac{1}{3}\sin 3\,\omega t + \frac{1}{5}\sin 5\,\omega t + \ldots\right) \quad (188)$$

Graphical application

Equations 184, 185 and 186 give the coefficients c, a_n and b_n in the Fourier series represented by equation 176 for any single-valued function. They can be applied, however, only when the function is known analytically—as, for example :—

from $\quad 0$ to t_1, $y = f_1(t)$

from $\quad t_1$ to t_2, $y = f_2(t)$

from $\quad t_2$ to $\frac{2\pi}{\omega}$, $y = f_3(t)$

Sometimes the function to be expressed as a Fourier series is known graphically but not analytically, and in such cases an approximate graphical method of analysis must be applied. Any number k of ordinates at intervals of $\frac{2\pi}{k}$ are drawn, and the heights $y_0, y_1, y_2, \ldots, y_{k-1}$ of each ordinate measured. The larger the number of ordinates drawn, the closer will the approximation be; the order n of the highest harmonic that can be calculated with reasonable accuracy by means of a k-ordinate analysis is given by :—

$$n = \frac{k-2}{2}$$

Equation 184 can then be rewritten as :—

$$c = \frac{1}{k}(y_0 + y_1 + y_2 + \dots + y_{k-1})$$

$$\therefore \qquad c = \frac{1}{k}\sum_{m=0}^{m=k-1} y_m \tag{189}$$

and equation 185 can be written as :—

$$\begin{aligned} a_n &= \frac{1}{\pi}\int_0^{2\pi} y \sin n\omega t \,.\, d(\omega t) \\ &= 2 \times \text{average ordinate of the curve } y \sin n\omega t \\ &\simeq \frac{2}{k} \times \text{sum of } k \text{ ordinates} \\ &= \frac{2}{k}\left(y_0 \sin\frac{2\pi \,.\, 0n}{k} + y_1 \sin\frac{2\pi \,.\, 1n}{k} + y_2 \sin\frac{2\pi \,.\, 2n}{k} \right. \\ &\qquad \left. + \dots + y_{k-1}\sin\frac{2\pi\,(k-1)\,n}{k}\right) \\ &= \frac{2}{k}\sum_{m=0}^{m=k-1} y_m \sin mn\frac{2\pi}{k} \end{aligned} \tag{190}$$

Similarly

$$b_n = \frac{2}{k}\sum_{m=0}^{m=k-1} y_m \cos mn\frac{2\pi}{k} \tag{191}$$

12-ordinate analysis.—

If twelve ordinates are taken, ($k = 12$), as in Fig. 68, they must

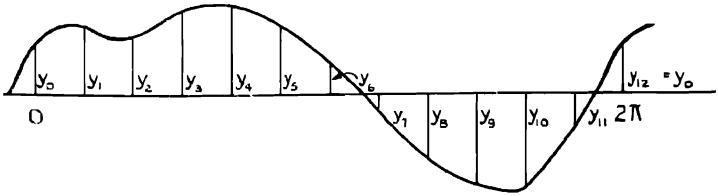
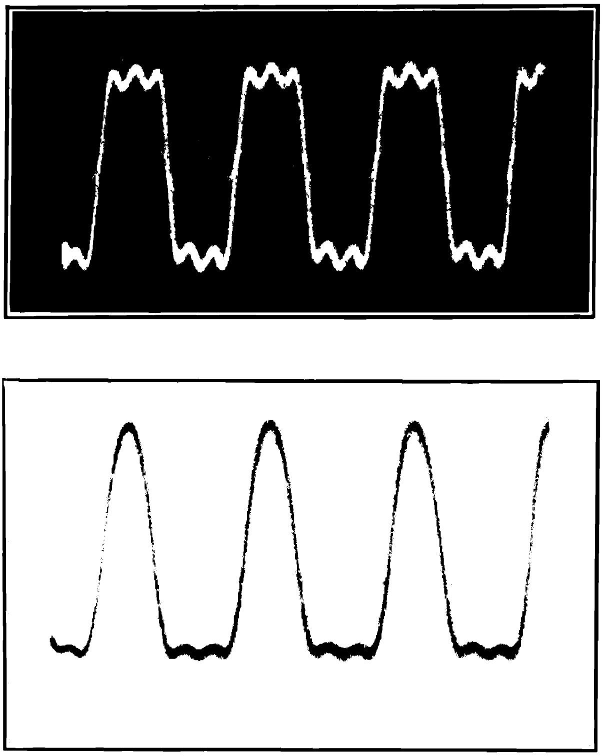


FIG. 68.—12-ordinate analysis.

spaced at $\frac{2\pi}{12}$, or 30°. Then :—

$$a_n = \frac{1}{6}\left\{ y_0 \sin 0n\frac{\pi}{6} + y_1 \sin 1n\frac{\pi}{6} + y_2 \sin 2n\frac{\pi}{6} + \dots \right.$$
$$\left. + y_{11} \sin 11n\frac{\pi}{6} \right\} \tag{192}$$



(a)

(b)

PLATE 2.—Oscillograms showing component sine waves and resultant wave forms.

and
$$b_n = \frac{1}{6}\left\{y_0 \cos 0n\frac{\pi}{6} + y_1 \cos n\frac{\pi}{6} + y_2 \cos 2n\frac{\pi}{6} + \ldots + y_{11} \cos 11n\frac{\pi}{6}\right\} \quad (193)$$

Hence the early coefficients in the series of equation 176 become :—

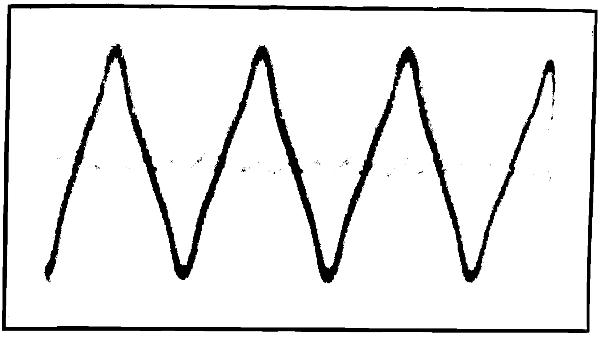
$$c = \frac{1}{12}\{y_0 + y_1 + y_2 + y_3 + y_4 + y_5 + y_6 + y_7 + y_8 + y_9 + y_{10} + y_{11}\} \quad (194)$$

$$a_1 = \frac{1}{6}\{(y_3 - y_9) + 0\cdot866\,(y_2 + y_4 - y_8 - y_{10}) + 0\cdot5\,(y_1 + y_5 - y_7 - y_{11})\} \quad (195)$$

$$a_2 = \frac{0\cdot866}{6}\{(y_1 + y_2 + y_7 + y_8) - (y_4 + y_5 + y_{10} + y_{11})\} \quad (196)$$

$$a_3 = \frac{1}{6}\{(y_1 + y_5 + y_9) - (y_3 + y_7 + y_{11})\} \quad (197)$$

(c)



(d)

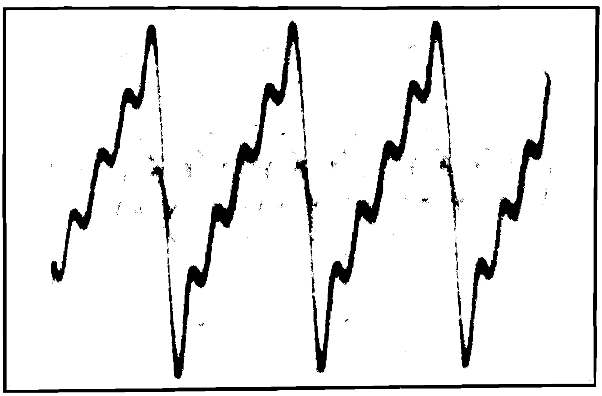


PLATE 2.—Oscillograms showing component sine waves and resultant wave forms

$$a_4 = \frac{0\cdot 866}{6}\{(y_1 + y_4 + y_7 + y_{10}) - (y_2 + y_5 + y_8 + y_{11})\} \tag{198}$$

$$a_5 = \frac{1}{6}\{(y_3 - y_9) + 0\cdot 866\,(y_8 + y_{10} - y_2 - y_4) + 0\cdot 5\,(y_1 + y_5 - y_7 - y_{11})\} \tag{199}$$

$$b_1 = \frac{1}{6}\{(y_0 - y_6) + 0\cdot 866\,(y_1 + y_{11} - y_5 - y_7) + 0\cdot 5\,(y_2 + y_{10} - y_4 - y_8)\} \tag{200}$$

$$b_2 = \frac{1}{6}\{(y_0 - y_3 + y_6 - y_9) + 0\cdot 5\,(y_1 + y_5 + y_7 + y_{11}) - 0\cdot 5\,(y_2 + y_4 + y_8 + y_{10})\} \tag{201}$$

$$b_3 = \frac{1}{6}\{(y_0 + y_4 + y_8) - (y_2 + y_6 + y_{10})\} \tag{202}$$

$$b_4 = \frac{1}{6}\{(y_0 + y_3 + y_6 + y_9) - 0\cdot 5\,(y_1 + y_2 + y_4 + y_5 + y_7 + y_8 + y_{10} + y_{11})\} \tag{203}$$

$$b_5 = \frac{1}{6}\{(y_0 - y_6) + 0\cdot 866\,(y_5 + y_7 - y_1 - y_{11}) + 0\cdot 5\,(y_2 + y_{10} - y_4 - y_8)\} \tag{204}$$

Example.—

Express as a Fourier series the curve shown in Fig. 69*a*. Applying a 12-ordinate analysis :—

$$\begin{array}{ll} y_0 = 3 & y_7 = -2 \\ y_1 = 2 & y_8 = -1 \\ y_2 = 1 & y_9 = 0 \\ y_3 = 0 & y_{10} = 1 \\ y_4 = -1 & y_{11} = 2 \\ y_5 = -2 & y_{12} = y_0 = 3 \\ y_6 = -3 & \end{array}$$

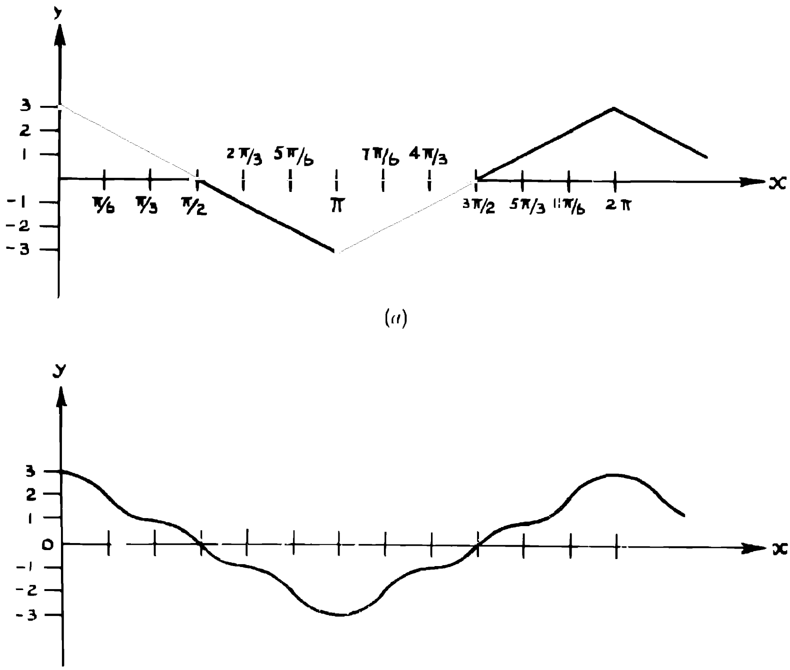


FIG. 69.—Example of Fourier analysis.

Then, from equations 194 to 204 :—

$$c = \frac{1}{12}\{3 + 2 + 1 + 0 - 1 - 2 - 3 - 2 - 1 + 0 + 1 + 2\} = 0$$

$$a_1 = \frac{1}{6}\{(0 - 0) + 0{\cdot}866\,(1 - 1 + 1 - 1) + 0{\cdot}5\,(2 - 2 + 2 - 2)\} = 0$$

$$a_2 = \frac{0{\cdot}866}{6}\{(2 + 1 - 2 - 1) - (-1 - 2 + 1 + 2)\} = 0$$

$$a_3 = \frac{1}{6}\{(2-2+0)-(0-2+2)\} = 0$$

$$a_4 = \frac{0.866}{6}\{(2-1-2+1)-(1-2-1+2)\} = 0$$

$$a_5 = \frac{1}{6}\{(0-0)+0.866(-1+1-1+1) + 0.5(2-2+2-2)\} = 0$$

$$b_1 = \frac{1}{6}\{(3+3)+0.866(2+2+2+2)+0.5(1+1+1+1)\} = \frac{1}{6}\{6+6.93+2\} = 2.49$$

$$b_2 = \frac{1}{6}\{(3-3)+0.5(2-2-2+2) - 0.5(1-1-1+1)\} = 0$$

$$b_3 = \frac{1}{6}\{(3-1-1)-(1-3+1)\} = 0.333$$

$$b_4 = \frac{1}{6}\{(3+0-3+0)-0.5(2+1-1-2-2-1+1+2)\} = 0$$

$$b_5 = \frac{1}{6}\{(3+3)+0.866(-2-2-2-2)+0.5(1+1+1+1)\} = \frac{1}{6}\{6-6.93+2\} = 0.178$$

Thus the curve is represented approximately by a series, the first three terms of which are (see Fig. 69*b*)

$$y = 2.49 \cos x + 0.333 \cos 3x + 0.178 \cos 5x + \ldots$$

Input-output curves

Many pieces of equipment used in communication engineering ve a non-linear characteristic, that is to say, the output is not irectly proportional to the input so that the waveform of the utput differs from that of the input.

The input-output characteristic of many items of equipment can e represented by a line PQ (see Fig. 70) such that, if any input, presented by a distance Oa along one axis Ox be projected on to t N the projection Ob of N on to Oy represents the output. PQ is a straight line the output will be identical in waveform the input, but if PQ is curved the output waveform will differ m the input waveform. This difference in waveform corresponds the generation of harmonics of the fundamental input frequency, t can be analysed by considering the output waveform produced n a sinusoidal input is applied.

Fig. 71 shows a sine wave so applied to a curve PQ that the jection of the peaks of the wave coincide with the ends P and Q he curve. The axis of the sine wave is extended to meet PQ at O, a line AOB drawn at right angles to this axis with $OA = OB$ plitude of sine wave. If 12 ordinates q_1 q_2 . . . are shown on

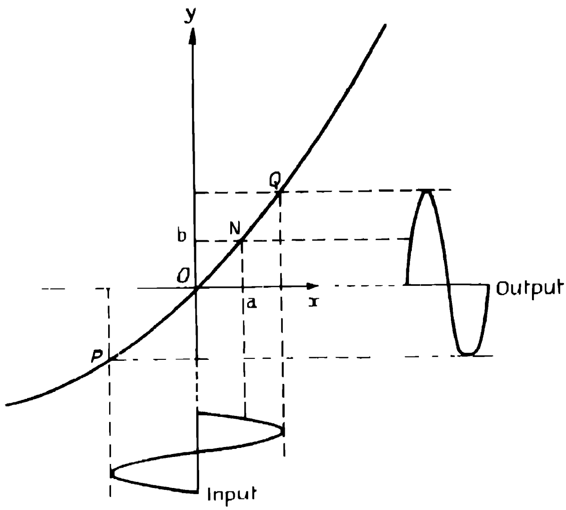


FIG. 70.—Typical input-output curve showing non-linearity.

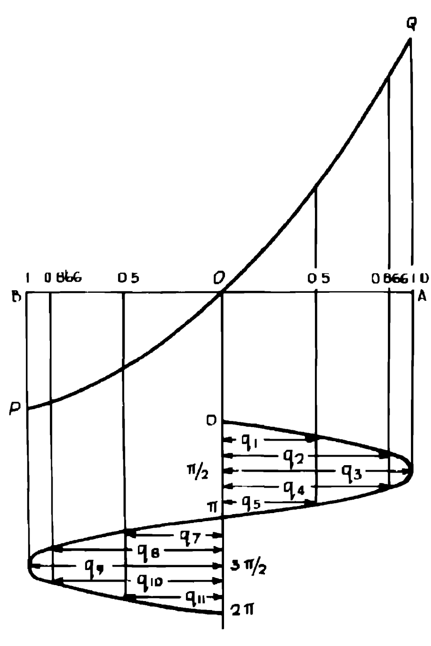


FIG. 71.—Application of 12-point analysis to input-output curve

the sine wave at intervals of $\frac{2\pi}{12}$, it can be seen that :—

$$\left.\begin{array}{llll} q_0 = q_6 & = q_{12} & = & 0 \\ q_1 = q_5 & = OA \sin 30^\circ & = & 0.5\,OA \\ q_2 = q_4 & = OA \sin 60^\circ & = & 0.866\,OA \\ q_3 & & = & OA \\ q_7 = q_{11} & = OA \sin 210^\circ & = & -0.5\,OA \\ q_8 = q_{10} & = OA \sin 270^\circ & = & -0.866\,OA \\ q_9 & & = & -OA \end{array}\right\} \quad (205)$$

An input-output curve PQ can thus be analysed by marking off, along the input axis distances equal to ± 0.5, ± 0.866, and ± 1 times the amplitude of the input under consideration, and by measuring the heights h_1, h_2, etc., of the curve at these points. If desired, the horizontal line AB may be taken not through O, but lower, so that all the heights h appear positive. the mean value term (c in equation 176) is then increased by h', where h' is the distance from O to AB, but the other terms in the series are unaffected.

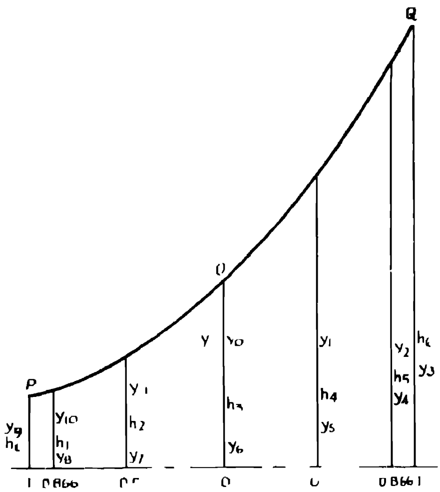


FIG. 72.—12 point analysis of input output curve

The coefficients of the terms in equation 176 can be found from equations 194 to 204. The notation of Fig. 72 gives :—

$$\left.\begin{array}{ll} y_0 = y_6 = y_{12} & = h_3 \\ y_1 = y_5 & = h_4 \\ y_2 = y_4 & = h_5 \\ y_3 & = h_6 \\ y_7 = y_{11} & = h_2 \\ y_8 = y_{10} & = h_1 \\ y_9 & = h_0 \end{array}\right\} \quad (206)$$

With this notation, these coefficients are :—

$$c = \frac{1}{12} \{(h_0 + h_6) + 2\,(h_1 + h_2 + h_3 + h_4 + h_5)\} \tag{207}$$

$$a_1 = \frac{1}{6} \{(h_4 + h_6) - (h_0 + h_2) + 1{\cdot}732\,(h_5 - h_1)\} \tag{208}$$

$$a_3 = \frac{1}{6} \{(h_0 - h_6) + 2\,(h_4 - h_2)\} \tag{209}$$

$$a_5 = \frac{1}{6} \{(h_6 - h_0) + (h_4 - h_2) + 1{\cdot}732\,(h_1 - h_5)\} \tag{210}$$

$$b_2 = \frac{1}{6} \{(h_2 + h_4 - h_5 - h_1) + 2h_3)\} \tag{211}$$

$$b_4 = \frac{1}{6} \{(h_0 + 2h_3 + h_6) - (h_1 + h_2 + h_4 + h_5)\} \tag{212}$$

$$a_2 = a_4 = b_1 \qquad b_3 = b_5 = 0 \tag{213}$$

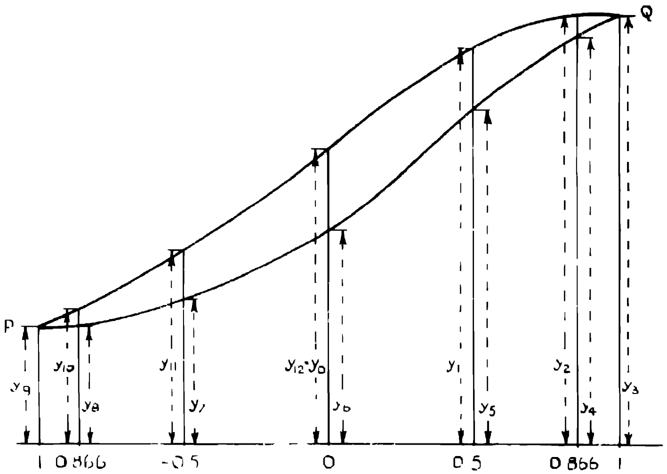
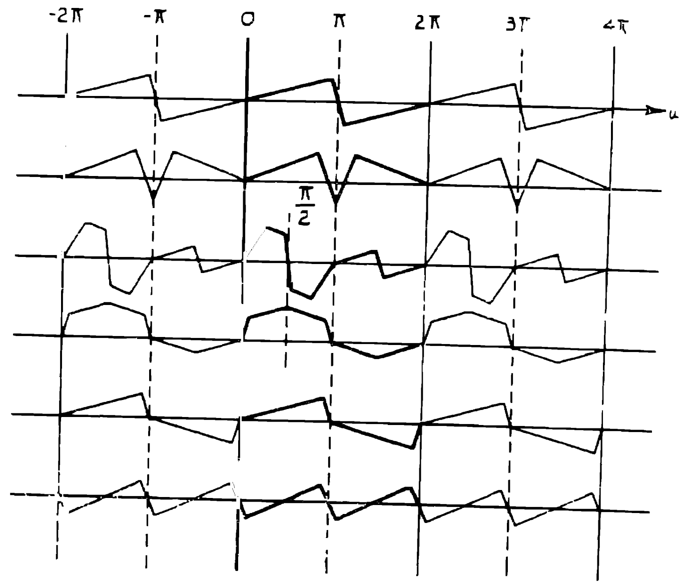


FIG. 73.—12-point analysis of input-output loop.

In certain cases, the input-output relationship must be represented not by a single line PQ, but by a closed loop. In such cases, equations 206 (and, therefore, 207 to 213) do not hold, but equations 194 to 204 may be applied, where y_1, y_2, *etc.*, have the meanings indicated in Fig. 73.

Particular cases of symmetry

In many cases, examination of the symmetrical properties of a curve may avoid the necessity for calculating the values of *all* the coefficients, since some of these may be seen, from inspection of the curve, to be zero. Fig. 74 gives the most important cases of symmetry.



Symmetry.	*Coefficients that cannot be present.* *Sine.*	*Cosine.*
(*a*) Skew-symmetrical about $\omega t = 0$ i.e. $f(\omega t) = -f(-\omega t)$	None	All
(*b*) Symmetrical about $\omega t = 0$ i.e. $f(\omega t) = -f(-\omega t)$	All	None
(*c*) Skew-symmetrical about $\omega t = \pm\frac{\pi}{2}$ i.e. $f(\omega t) = -f(\pi - \omega t)$	Odd	Even
(*d*) Symmetrical about $\omega t = \pm\frac{\pi}{2}$ i.e. $f(\omega t) = +f(\pi - \omega t)$	Even	Odd
(*e*) Both half-cycles of same form, *but with* $f(\omega t) = -f(\pi + \omega t)$	Even	Even
(*f*) Both half-cycles of same form; *but with* $f(\omega t) = +f(\pi + \omega t)$ [This is really a double frequency waveform, since the fundamental frequency $n = 1$ is excluded.]	Odd	Odd

FIG 74 —Particular cases of symmetry in Fourier analysis.

A curve is said to be *symmetrical* about a line $\omega t = k$ if its shape to the left of that line is a " mirror image " of its shape to the right of it ; while it is *skew-symmetrical* about that line if its shape to the left of it is the same as that to the right, but inverted about the horizontal (ωt) axis. If the curve is symmetrical about $y = 0$ (as it is, of necessity, in cases (*a*), (*c*) and (*e*)), then the constant a_0 is zero. In other cases it may be necessary to subtract the component $y = a_0$ before the above conditions of skew-symmetry can be realised.

If a curve exhibits symmetry about two points, each of these will impose a limitation on the coefficients that may be present. The square waveform (Fig. 67) exhibits symmetry about $\omega t = \pm\frac{\pi}{2}$ and skew-symmetry about $\omega t = 0$. Its series therefore contains only *odd sine terms*.

Analysis of typical waveforms

(*a*) *Saw-toothed.*—

$$y = \frac{2d}{\pi}\left\{\sin x - \frac{1}{2}\sin 2x + \frac{1}{3}\sin 3x - \ldots\right\}$$

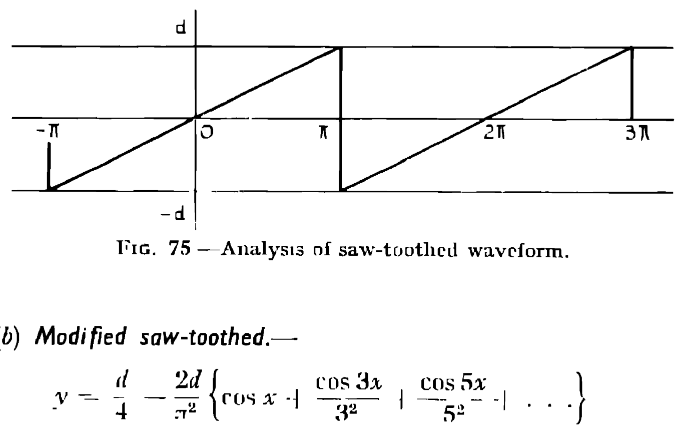


Fig. 75.—Analysis of saw-toothed waveform.

(*b*) *Modified saw-toothed.*—

$$y = \frac{d}{4} - \frac{2d}{\pi^2}\left\{\cos x + \frac{\cos 3x}{3^2} + \frac{\cos 5x}{5^2} + \ldots\right\}$$
$$+ \frac{d}{\pi}\left\{\sin x - \frac{\sin 2x}{2} + \frac{\sin 3x}{3} - \ldots\right\}$$

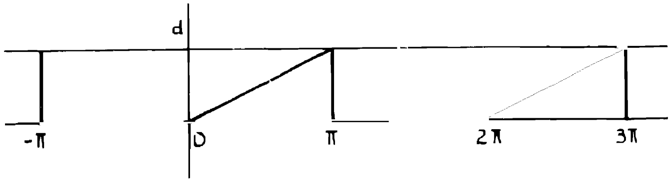


Fig. 76.—Analysis of modified saw-toothed waveform.

(*c*) *Triangular.*—

$$y = \frac{8d}{\pi^2}\left\{\cos x + \frac{\cos 3x}{3^2} + \frac{\cos 5x}{5^2} + \ldots\right\}$$

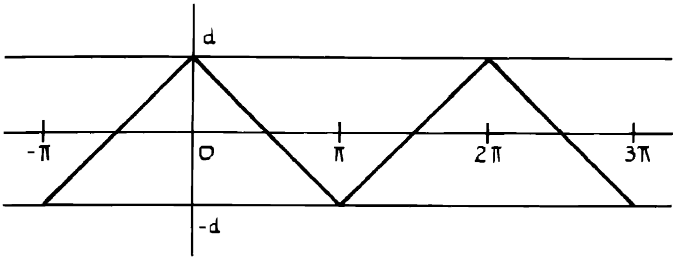


FIG. 77.—Analysis of triangular waveform.

(*d*) *Half-wave rectifier output.*—

$$y = \frac{2d}{\pi}\left\{\frac{1}{2} + \frac{\pi}{4}\sin x - \frac{1}{1.3}\cos 2x - \frac{1}{3.5}\cos 4x - \frac{1}{5.7}\cos 6x - \ldots\right\}$$

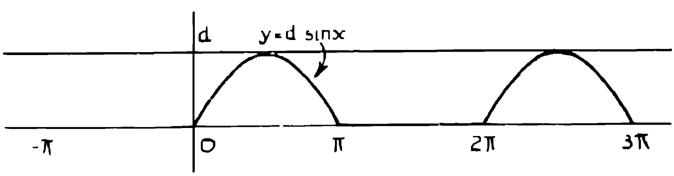


FIG. 78.—Analysis of half-wave rectifier output.

(*e*) *Full-wave rectifier output.*—

$$y = \frac{4d}{\pi}\left\{\frac{1}{2} - \frac{1}{1.3}\cos 2x - \frac{1}{3.5}\cos 4x - \frac{1}{5.7}\cos 6x - \ldots\right\}$$

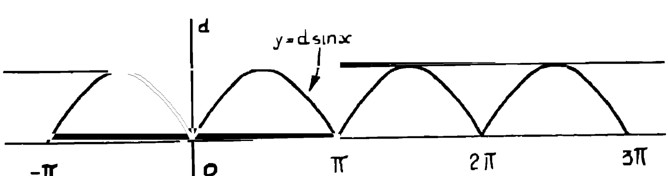


FIG. 79.—Analysis of full-wave rectifier output.

CHAPTER 3

DIRECT CURRENTS

This chapter summarises some of the important principles of elementary electricity and magnetism that are applicable to line communication, and is intended for revision

Electric currents are intimately concerned with the structure of matter, and the answer to the question "What is electricity?" cannot be given without some knowledge of this structure. A very brief description of the molecules and atoms from which all matter is built up will therefore be given.

The smallest particle of any substance that can exist independently and still retain the characteristics of that substance is called a "molecule" of that substance. Each molecule is built up from a number of chemical elements, the smallest particle of each element being known as an atom of that element. Table IV gives a list of most of the known elements. The existence of molecules and atoms as the fundamental particles of matter has been accepted since the eighteenth century, but it was not until the present century that a satisfactory picture of the structure of the atom was provided.

Atomic structure

The atom of any particular element is now believed to consist of a positively-charged "nucleus" surrounded by a number of "planetary electrons". These planetary electrons behave as extremely small negative charges rotating in orbits of various sizes around the central nucleus which constitutes almost the entire mass of the atom The epithets "positively-" and "negatively-charged" were originally used to express the fact that two similarly "charged" bodies (*e g*, both positive) repel each other, whereas two bodies of opposite charge (one positive, one negative) attract each other. The electron is considered to be the fundamental electric charge, a negatively charged body being one having a surplus of electrons, and a positively charged body being one having a deficit of them.

The charge of the electron is extremely small; the practical unit of electrical charge or "quantity of electricity", the coulomb, is equivalent to a definite, very large, number of electrons (actually about $6 \cdot 28 \times 10^{18}$).

TABLE IV

Atomic weights and numbers

Element	Symbol	Atomic number	Atomic weight	Element	Symbol	Atomic number	Atomic weight
Actinium	Ac	89	c.227	Neodymium	Nd	60	144·27
Aluminium	Al	13	26·97	Neon	Ne	10	20·18
Antimony	Sb	51	121·76	Nickel	Ni	28	58·69
Argon	A	18	39·94	Niobium	Nb	41	92·91
Arsenic	As	33	74·91	Nitrogen	N	7	14·01
Barium	Ba	56	137·36	Osmium	Os	76	190·2
Beryllium	Be	4	9·02	Oxygen	O	8	16·00
Bismuth	Bi	83	209·00	Palladium	Pd	46	106·7
Boron	B	5	10·82	Phosphorus	P	15	31·02
Bromine	Br	35	79·92	Platinum	Pt	78	195·23
Cadmium	Cd	48	112·41	Polonium	Po	84	210
Caesium	Cs	55	132·91	Potassium	K	19	39·10
Calcium	Ca	20	40·08	Praeseodymium	Pr	59	140·92
Carbon	C	6	12·00	Protoactinium	Pa	91	231
Cerium	Ce	58	140·13	Radium	Ra	88	226·05
Chlorine	Cl	17	35·46	Radon	Rn	86	222
Chromium	Cr	24	52·01	Rhenium	Re	75	186·31
Cobalt	Co	27	58·94	Rhodium	Rh	45	102·91
Copper	Cu	29	63·57	Rubidium	Rb	37	85·48
Dysprosium	Dy	66	162·46	Ruthenium	Ru	44	101·7
Erbium	Er	68	167·20	Samarium	Sm	62	150·43
Europium	Eu	63	152·0	Scandium	Sc	21	45·10
Fluorine	F	9	19·00	Selenium	Se	34	78·96
Gadolinium	Gd	64	156·9	Silicon	Si	14	28·06
Gallium	Ga	31	69·72	Silver	Ag	47	107·88
Germanium	Ge	32	72·60	Sodium	Na	11	23·00
Gold	Au	79	197·2	Strontium	Sr	38	87·63
Hafnium	Hf	72	178·6	Sulphur	S	16	32·06
Helium	He	2	4·00	Tantalum	Ta	73	180·88
Holmium	Ho	67	163·5	Tellurium	Te	52	127·61
Hydrogen	H	1	1·008	Terbium	Tb	65	159·2
Indium	In	49	114·76	Thallium	Tl	81	204·39
Iodine	I	53	126·92	Thorium	Th	90	232·12
Iridium	Ir	77	193·1	Thulium	Tm	69	169·4
Iron	Fe	26	55·84	Tin	Sn	50	118·70
Krypton	Kr	36	83·7	Titanium	Ti	22	47·90
Lanthanum	La	57	138·92	Tungsten	W	74	183·92
Lead	Pb	82	207·22	Uranium	U	92	238·07
Lithium	Li	3	6·94	Vanadium	V	23	50·95
Lutecium	Lu	71	175·0	Xenon	Xe	54	131·3
Magnesium	Mg	12	24·32	Ytterbium	Yb	70	173·04
Manganese	Mn	25	54·93	Yttrium	Y	39	88·92
Mercury	Hg	80	200·61	Zinc	Zn	30	65·38
Molybdenum	Mo	42	95·95	Zirconium	Zr	40	91·22

The number of planetary electrons in each atom varies with the element; hydrogen, the lightest, has only one; helium has two; lithium three; and so on up to uranium, which has 92. Elements heavier than uranium are known to exist, and are classified under the general heading of "trans-uranic" elements. Thus according to

the element considered, the atom will contain between 1 and 90-odd units of negative charge. A complete atom, however, is electrically neutral, these negative charges being counteracted by equal positive charges in the nucleus. The number of such charges is known as the "atomic number" of the element (*see* Table IV).

The nucleus contains a number of particles called "protons"; a proton has a weight roughly 1,840 times that of the electron, and has a charge equal to that of the electron but opposite in sign. But in all atoms (except hydrogen) the weight of the planetary electrons together with that of the (equal number of) protons in the nucleus does not account for the whole weight of the atom. The weight of an atom (*i.e.*, atomic weight) is normally measured with reference to the atom of oxygen, which is assumed to have a weight of 16·00. The original theory assumed that the additional mass was due to equal number of protons and electrons added to the nucleus; the charge of each additional electron would balance that of each additional proton, and the net charge on the atom would still be zero. This theory was modified, however, in 1932 by the discovery of the "neutron"; this is a particle of mass roughly equal to that of the proton, but with zero charge. On the new theory, each additional proton with its associated electron is believed to be replaced by a neutron.

In the ordinary state, many elements consist of a mixture of "isotopes". These are atoms having the same atomic number but different atomic weights, and they are caused by different numbers of neutrons in the nucleus. The various isotopes of an element have identical chemical, but differing physical, properties.

Thus an atom is now assumed to consist of a nucleus of protons and neutrons, with planetary electrons (equal in number to the protons in the nucleus, and to the atomic number of the element) rotating in orbits around it.

Other particles are also known to exist, such as the positron and the mesotron, but these are less important.

It has been seen that an atom is normally electrically neutral. When an atom or molecule contains a surplus or deficiency of electrons, it is said to have been ionised. If an excess of electrons occurs, the atom or molecule will exhibit the properties of a negatively-charged particle; in such a state it is known as a "negative ion". If, however, a deficiency of electrons occurs, the atom or molecule will exhibit the properties of a positively-charged particle, and is known as a "positive ion".

The electric current

It has been seen that the electrons rotate in orbits around the nucleus, and these orbits are maintained by the electrostatic attraction between the electrons and the nucleus. Electrons in the outer orbits, being further from the nucleus, are held more loosely; in fact, in certain substances, such as metals, transfer of

these outer electrons between adjacent atoms is continually taking place. Such substances are called " conductors ". If electrons are removed, by external means, from one end of a conductor, these loosely held electrons will be attracted towards that end, and a resultant motion of electrons will ensue. This flow of electrons constitutes an electric current. Substances in which the electrons are tightly bound to their parent nuclei will not permit such a flow of electrons; these substances are called " insulators ".

To take a specific case, a simple cell consists of a copper rod or " electrode " and a zinc electrode inserted in dilute sulphuric acid. When the electrodes are connected externally by a metallic wire, it is found that an electric current flows due to chemical action

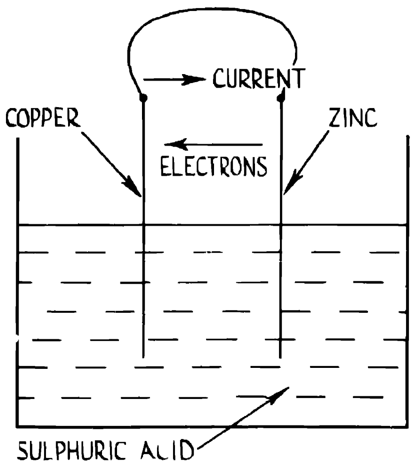


FIG. 80.—Simple cell.

between the electrodes and the acid. Consider the current flowing through this wire. It was originally assumed that " electricity " flowed from the copper electrode to the zinc; for this reason, the copper electrode was called the " positive ", and the zinc the " negative ". Thus " conventional current " flows from the positive electrode (copper) to the negative (zinc), although it is now known that what is actually happening is that a movement of the loosely held orbital electrons in the wire is taking place in the direction of zinc to copper. Thus one may say that electrons are flowing from the zinc to the copper. It is important to note that nearly all the laws of electricity are worded on the basis of " conventional current flow ", as they were stated before electrons had been discovered.

The practical unit of current is the " ampere ", which is equal to *a rate of movement of electric charge of one coulomb per second* (a flow of $6 \cdot 28 \times 10^{18}$ electrons per second).

More exactly, any movement of electric charge constitutes an electric current. In addition to a movement of electrons, a current

may be produced by a movement of positively and negatively charged ions

By convention, the direction of flow of an electric current is determined by considering the flow to be due to positively charged particles Thus, although a flow of electricity from point A to point B may be due to positively charged ions moving from A to B, it may also be due to negatively charged particles such as electrons or negative ions flowing from B to A (*see* Fig 81)

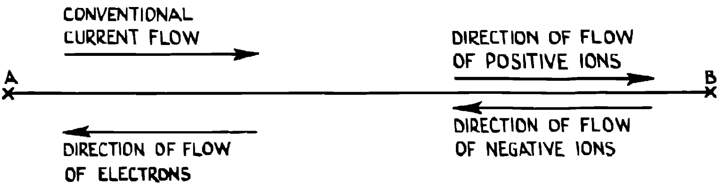


FIG 81 —Direction of flow of electrons or ions to give flow of current.

The passage of electricity may occur —

(*a*) through a conductor such as a metal due to the movement of the loosely held outer electrons of the atoms ,

(*b*) through a vacuum or gas due to the movement of electrons ,

(*c*) through a gas due to a movement of the ionised gas molecules ,

(*d*) through a liquid due to the ionisation of certain molecules particularly those of acids and salts in solution, and the movement of the resulting ions

Electro-motive force and potential difference

It is convenient to regard an ' electric pressure as being set up between the electrodes of a cell , this pressure is known as ' electromotive force " (EMF). It is dependent solely upon the chemical constitution of the cell and it exists even when the connecting wire is removed *i e* when no current is flowing In the example given above, the copper electrode is said to be at a " higher potential " than the zinc, so that a "potential difference ' (PD) exists between the two ends of the wire joining the electrodes It is this PD which causes the current to flow through the wire from the point of higher to the point of lower potential, and it is the EMF of the cell which produces this PD The practical unit used both for EMF and for PD is the ' volt ", which is $\frac{1}{1 \cdot 0186}$ of the EMF of a standard Weston cadmium cell at a temperature of 20° C A more rigid definition of this and other electrical units is given on page 131

Conductors and insulators

It has been seen that conductors are those substances that permit the movement of electrons from atom to atom through them

when a potential difference is applied; the ease with which electrons can be removed from their orbits by such a PD varies as from substance to substance. An insulator or non-conductor, on the other hand, is a substance in which the outer orbital electrons are tightly held to the atomic nuclei and will not break away on the application of a potential difference. If a PD is maintained between the ends of an insulator, the orbital electrons will be pulled over towards the point of higher potential, and the result will be a distortion of the electron orbits. This hypothesis will be used later to explain certain phenomena in connection with dielectric materials.

RESISTANCE

The distinction between conductors and insulators is not well defined, and there are many substances that may be regarded either as poor conductors or as poor insulators. In fact, all substances offer some "resistance" to the movement of electrons through them, and the magnitude of this resistance varies from a very low value in the case of a good conductor, through intermediate values for poor conductors and poor insulators, to a high value in the case of good insulators. The precise meaning of electrical resistance, and the unit in which it is measured, are explained in the next paragraph.

Ohm's Law

As a result of practical measurement, it was found by Ohm that :—

In any conductor, the ratio between the potential difference across it and the current flowing through it is a constant, provided that the physical conditions of the conductor, such as temperature, remain unchanged.*

This constant is termed the "resistance" of the conductor, and the practical unit is the "Ohm". This is defined as follows :—

If a PD of 1 volt across a conductor causes a current of 1 ampere to flow through it, then the resistance of that conductor is 1 ohm.

Using this definition, it follows that :—

$$\frac{E}{I} = R \tag{1}$$

where E is potential difference in volts,

I is the current in amperes,

and R is the resistance in ohms.

*There are a few exceptions which are dealt with in Chapter 6, but which need not be considered at this stage. The statement made here is true for all pure metals and alloys.

It is this mathematical expression which is usually known as "Ohm's Law", and it is often quoted in the equivalent forms:

$$E = IR \qquad (2)$$

or

$$I = \frac{E}{R} \qquad (3)$$

Resistances in series

With resistances in series the current will be the same in each. Applying Ohm's law to each in turn —

$$E_1 = IR_1$$
$$E_2 = IR_2$$
$$E_3 = IR_3$$

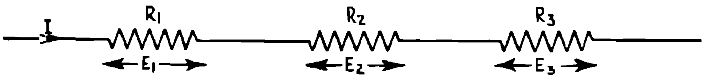


Fig. 82.—Resistances in series.

∴ The total potential difference E is

$$E = E_1 + E_2 + E_3$$
$$= IR_1 + IR_2 + IR_3$$

But if R is resultant resistance then by Ohm's law —

$$E = IR$$
$$IR = IR_1 + IR_2 + IR_3$$

or

$$R = R_1 + R_2 + R_3 \qquad (4)$$

Resistances in parallel

The potential difference across all three resistances is the same, *viz.*, the P.D. E volts between A and B. The total current I is given by —

$$I = I_1 + I_2 + I_3$$

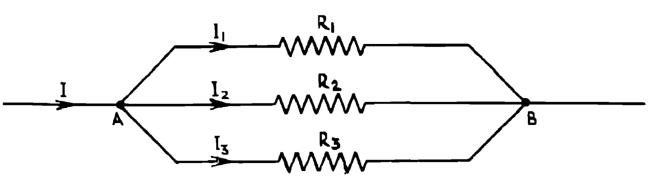


Fig. 83.—Resistances in parallel.

Applying Ohm's law to each resistance in turn —

$$I = \frac{E}{R_1} + \frac{E}{R_2} + \frac{E}{R_3}$$

But if R is the resultant resistance then by Ohm's law —

$$I = \frac{E}{R}$$

$$\therefore \frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} \tag{5}$$

When two resistances are in parallel it will be noted that :—

since $$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2}$$

$$\therefore \quad R = \frac{R_1 R_2}{R_1 + R_2} \tag{6}$$

The reciprocal of resistance is often termed the "conductance", and is measured in mhos (sometimes called reciprocal ohms); that is :—

$$G \text{ (in mhos)} = \frac{1}{R \text{ (in ohms)}}$$

Then for conductances in parallel, the total conductance G is :—

$$G = G_1 + G_2 + G_3 + \ldots \tag{7}$$

Specific resistance

The resistance of a conductor depends both on its dimensions and on the material of which it is made, and it is desirable to compare the resistive properties of materials in some way that is independent of the dimensions of the conductor. The "specific resistance" of a material (also known as resistivity) is defined as *the resistance between the opposite faces of a 1 cm cube of the material,* and is measured in "ohms per cm cube". Then the resistance R of a conductor of that material is given by

$$R = \frac{\rho l}{A} \tag{8}$$

where l is the length of the conductor (in cm.),
A is the area of the conductor (in sq. cm.),
and ρ is the specific resistance of the material.

The reciprocal of specific resistance or resistivity is specific conductance or conductivity, which is measured in mhos per cm. cube.

Temperature Coefficient of resistance

As a general rule the specific resistance of a metallic conductor increases with rise in temperature. The following equation gives the relationship between the resistance, R_t at $t°C$ and the resistance R_0 at 0° C., for a very wide range of temperature.

$$R_t = R_0 (1 + \alpha t + \beta t^2) \tag{9}$$

where α and β are constants for the metal concerned. Over a moderate range of temperature, say 0° to 100° C., the constant β, which is very small, has negligible effect, and the equation :—

$$R_t = R_0 (1 + \alpha t) \tag{10}$$

is sufficiently accurate. α is called the temperature coefficient of resistance Table V gives average values of ρ and α.

TABLE V

Specific resistances and temperature coefficients

Metal	Specific resistance ρ (ohms per cm cube)	Temperature coefficient of resistance α
Copper	1.6×10^{-6}	0.004
Iron	9.8×10^{-6}	0.006
Manganin	44.0×10^{-6}	0.00002

Kirchhoff's laws

These two laws are of universal application in the treatment of electrical networks For the purpose of this section a network will be defined as any number of resistances and batteries connected together to form an electrical circuit

Law 1 *The algebraic sum of currents meeting at any point in a network is zero*

Law 2 *The algebraic sum of the products of current and resistance in each conductor of a closed circuit is equal to the algebraic sum of the EMFs in that circuit*

The first law merely states that the current entering a point is equal to the current leaving it This is equivalent to saying that there is no accumulation of charge at the point The second law can be verified by application of Ohm's law the term algebraic merely signifying that if clockwise currents or EMFs are considered as positive then anti-clockwise ones must be treated as negative The law therefore states that the algebraic sum of EMFs and PDs in any closed circuit is zero

Example —

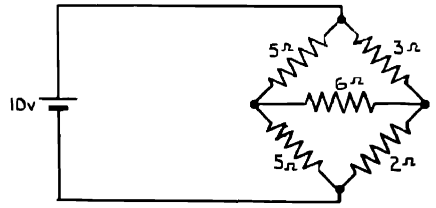


FIG 84 (*a*)

The network shown in Fig 84*a* has a potential of 10v applied to it from a battery that has no internal resistance Find the current in the centre arm (6 ohms)

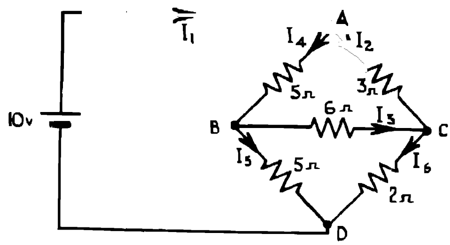


FIG 84 (*b*)

Fig 84*b* shows the circuit with the currents indicated as I_1 to I_6

But by Kirchhoff's first law the currents leaving the point A are equal to the currents entering,

$$\therefore \qquad I_4 = I_1 - I_2$$

Similarly at B $\qquad I_5 = I_4 - I_3 = I_1 - I_2 - I_3$

and at C $\qquad I_6 = I_2 + I_3$

To check these results, the current leaving D

$$\begin{aligned} &I_5 + I_6 \\ &= I_1 - I_2 - I_3 + I_2 + I_3 \\ &= I_1 \end{aligned}$$

This must be correct, since I_1 was originally the current in this circuit.

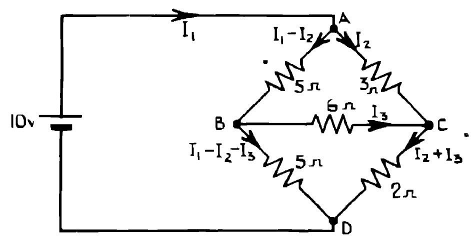


FIG 84 (*c*)

Using these results (*see* Fig 84*c*), the unknown quantities are now reduced to three—namely I_1, I_2 and I_3—and it will be necessary to obtain three equations by application of Kirchhoff's second law to the network At least one of these equations must contain the battery EMF

Consider *ACB* This is a closed network, and therefore Kirchhoff's second law may be applied to it —

$$3I_2 - 6I_3 - 5\,(I_1 - I_2) = 0,$$

there being no applied EMF.

On simplifying this,

$$5I_1 - 8I_2 + 6I_3 = 0 \qquad \text{(i)}$$

Consider BCD :—

$$6I_3 + 2(I_2 + I_3) - 5(I_1 - I_2 - I_3) = 0$$

Simplifying, $$5I_1 - 7I_2 - 13I_3 = 0 \qquad \text{(ii)}$$

Consider outer network, ACD and battery :—

$$3I_2 + 2(I_2 + I_3) = 10$$

Simplifying, $$5I_2 + 2I_3 = 10 \qquad \text{(iii)}$$

From equations (i) and (ii) —

$$I_2 - 19I_3 = 0$$
$$I_2 = 19I_3$$

Substitute for I_2 in equation (iii)

$$5(19I_3) + 2I_3 = 10$$
$$I_3 = \frac{10}{97}$$
$$= 0.103 \text{ amps} \quad \textit{Ans}$$

Measurement of resistance

There are three simple methods of measuring resistance, namely, by voltmeter and ammeter, by substitution, and by Wheatstone's bridge.

1. *Voltmeter and ammeter*.—A suitable voltage is applied to either of the networks shown in Fig 85*a* and 85*b*. From the readings of the two meters, the value of R can be calculated by

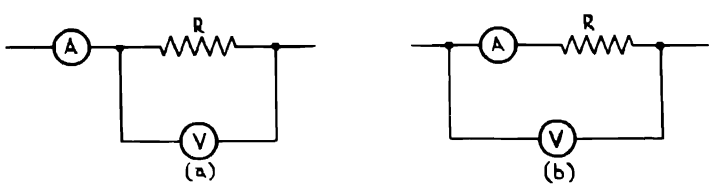


FIG 85 – Measurement of resistance by voltmeter and ammeter method.

Ohm's law. Errors are introduced in both methods, since in (*a*) the ammeter reading includes the current through the voltmeter, and in (*b*) the voltmeter reading includes the voltage drop across the ammeter. If the resistances of the meters are known, correction can, however, be made. The resistance of a voltmeter is usually known more accurately than that of an ammeter, and also it is usually high compared with the resistance to be measured; method (*a*) is therefore preferable in most cases.

2. *By substitution.*—Connect the resistance R to be measured in series with a galvanometer and battery. Note the deflection obtained. Replace R by a calibrated resistance r and vary it until

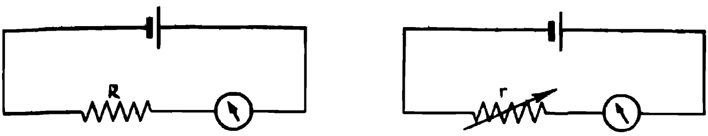


Fig 86 —Measurement of resistance by substitution

the galvanometer gives the same deflection as before, the value of r is then equal to R

Wheatstone's bridge

The bridge method for comparing and measuring resistances is very widely used and though it will be considered here only from the DC aspect, its application is of great importance in AC work as well

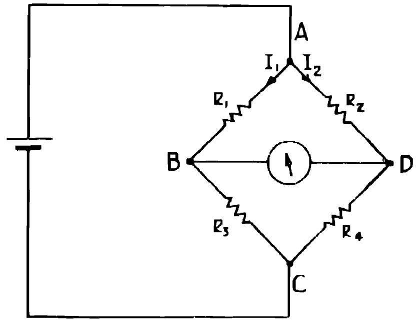


Fig 87 Measurement of resistance by Wheatstone's bridge

Consider the condition of the network illustrated that will give no deflection in the galvanometer i e no current in BD This means that B and D are at the same potential or in other words, the PD across AB = the PD across AD

$$\therefore \quad I_1R_1 = I_2R_2$$

$$\therefore \quad \frac{I_1}{I_2} = \frac{R_2}{R_1}$$

Since there is no current in BD then by applying Kirchhoff's first law to the point B the current in BC must be I_1 Similarly, the current in DC is I_2

As before, the PD across BC = the PD across DC

$$\therefore \quad I_1R_3 = I_2R_4$$

$$\therefore \quad \frac{I_1}{I_2} = \frac{R_4}{R_3}$$

$$\text{Hence} \quad \frac{R_2}{R_1} = \frac{R_4}{R_3}$$

This gives the condition that the bridge shall be balanced.

Therefore, if this condition can be reached, a simple relationship is set up between all four resistances; if the ratio $\frac{R_2}{R_1}$ and the value of R_3 are known, it is easy to calculate R_4. This system is used in the metre bridge and in the Post-Office box. In the former, the known resistance is fixed and the ratio is varied; in the latter, the ratio is fixed and the known resistance is varied. Each gives a high degree of accuracy.

ELECTRICAL UNITS

In addition to the "practical" system of units, which is used for everyday electrical work and of which some examples have already been given, there exist three other systems of units. Two of these, the "electrostatic" and the "electromagnetic", are "absolute" systems based on theoretical considerations; while not used for practical measurements, they are none the less important to a study of electrical theory. The other system of units is the "international", which forms a standard against which the practical units can be compared.

Electrostatic units (ESU)

The "electrostatic unit of charge" is defined as *that charge which, when placed 1 cm. distant from an exactly similar charge in a vacuum, repels it with a force of 1 dyne.* Measurement of charge is simply a measurement of the number of electrons. If the electron had been discovered before the ESU of charge was defined, the electron might equally well have been made the basic unit of charge in theoretical considerations.

$$1 \text{ ESU of charge} \simeq 2 \cdot 09 \times 10^9 \text{ electrons.}$$

Electric current is measured as a rate of flow of charge (or electrons). The "electrostatic unit of current" is defined as *a rate of flow of one ESU of charge per second.*

Electric potential is a measure of the potential energy at a point due to its position in an electrostatic field. The actual value of the potential at a point, like all potential energies, is purely relative, and only the *potential difference* between two points is completely determinable. The "electrostatic unit of potential" is defined as *the potential difference that exists between two points when the work done in taking one ESU of charge from one point to the other is one erg [1 erg = 1 dyne-cm.].*

Electromagnetic units (EMU)

For completeness, the principal electromagnetic units will now be defined, although these will not be referred to again until later in the chapter. The basis of the electromagnetic system of units is the force between magnetic poles, just as the basis of the electrostatic system is the force between electric charges.

Current is measured in terms of the electromagnetic field it produces. The "electromagnetic unit of current" is defined as *that current which, when flowing in a circular loop of wire of radius* 1 *cm., produces at the centre a magnetic field of 2π dynes per unit pole.*

The "electromagnetic unit of electric charge" is defined as *that charge which is transferred past a point of a conductor when a current of one EMU flows for one second.*

The "electromagnetic unit of potential" is defined as *that potential difference which exists between two points when the work done in taking one EMU of charge from one point to the other is one erg.*

Practical units

Neither the electrostatic system nor the electromagnetic system of units is completely satisfactory as a means of expressing practical electrical measurements, since certain of these units are inconveniently large, others much too small. The fundamental practical units are rigidly defined as follows :-

1 Ampere $= 10^{-1}$ EMU of current [$\simeq 3 \times 10^{9}$ ESU of current] (11)

1 Volt $= 10^{+8}$ EMU of potential [$\simeq \frac{1}{300}$ ESU of potential] (12)

1 Coulomb $= 10^{-1}$ EMU of charge [$\simeq 3 \times 10^{9}$ ESU of charge] (13)

These definitions were resolved in 1908 by an International Conference, which decided that the fundamental practical units should be based on the electromagnetic rather than on the electrostatic system. The connections quoted above between the practical units and the ESU are sufficiently accurate for almost all purposes.

International units

The international units were defined at the conference previously mentioned, in terms of the measurable physical quantities of mass, distance, and time. They are used as a basis of comparison and legislation. At the same time, their values correspond very closely with the fundamental practical units.

The "international ohm" is defined as *the resistance offered to an unvarying electric current by a column of mercury at 0° C., 14·4521 grammes in mass, of a constant cross-section, and of length 106·300 cms.*

The advantage of such a unit lies in the fact that it has a definite physical interpretation.

The *international ampere* is defined in terms of an electrochemical phenomenon, namely that if two electrodes are immersed in a solution of silver nitrate and a potential difference is maintained between them, a current will flow and silver will be deposited on the negative (lower potential) electrode. The rate of deposit of

silver is proportional to the current flowing. The definition is framed as follows :—

The " international ampere " is *that unvarying electric current which, when passed through an aqueous solution of silver nitrate, in accordance with an authorised specification, deposits silver at the rate of 0·0011180 grammes per second.*

The " international volt " is *that potential difference which must be applied across a conductor whose resistance is one international ohm in order to produce a current of one international ampere.*

POWER

Consider a conductor having a potential difference of e ESU between its two ends The statement that the PD has this value is, by definition, equivalent to saying that the work done in passing one ESU of charge from one end of the conductor to the other is e ergs. The work done in passing a charge of q ESU will therefore be qe ergs But if this process takes a time t secs, then

$$q = it \text{ where } i \text{ is the current in ESU}$$

$$\therefore \text{ Work done} = eit \text{ ergs.} \tag{14}$$

Now let the PD be E volts and the current I amps Then for all practical purposes $e = \frac{E}{300}$ and $i = 3 \times 10^9 I$

$$\therefore \text{ The work done} = EIt \times 10^7 \text{ ergs} \tag{15}$$

But 10^7 ergs = 1 joule (practical unit of work)
and work done = EIt joules

Power is the rate of doing work, *i.e.* :—

$$\text{Power} = EI \text{ joules per second} \tag{16}$$

The unit of power is called the " watt ".

1 watt	= 1 joule/second	(17)
1 Kilowatt	= 1,000 joules/second	(18)
1 kW-hour	= 1,000 × 3,600 joules	(19)
1 Horse power	= 746 watts	(20)
	= 550 ft.-lb./sec.	(21)
4·18 joules	= 1 calorie of heat energy.	(22)

The formula for power may be adapted as required, by using Ohm's law.

i.e.,
$$\text{Power} = E \times I \text{ watts} \tag{23}$$

$$= \frac{E^2}{R} \text{ watts} \tag{24}$$

$$= I^2R \text{ watts} \tag{25}$$

Maximum power transfer theorem (DC case)

Consider a battery or generator of EMF E and internal resistance r, supplying current to a load R.

By Ohm's law :—

$$I = \frac{E}{R + r}$$

Therefore the power P supplied to the load R is

$$P = I^2R$$

$$= E^2 \frac{R}{(R + r)^2}$$

It is required to find the value of R that will enable maximum power to be taken from the generator. Therefore in the expression

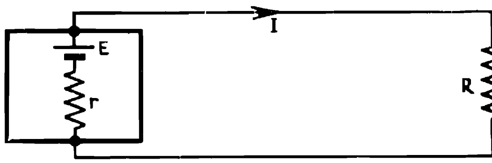


FIG. 88.—Maximum power transfer (DC case).

above, E and r will be considered constant and the expression differentiated with respect to R :—

$$P = E^2 \frac{R}{(R + r)^2}$$

$$\frac{dP}{dR} = E^2 \left[\frac{(R + r)^2 - 2R(R + r)}{(R + r)^4} \right]$$

$$= E^2 \left[\frac{r - R}{(R + r)^3} \right]$$

For a maximum or minimum value $\frac{dP}{dR} = 0$, $\therefore R = r$. (26)

It may be verified by a second differentiation that $R = r$ gives a *maximum* value of P. This therefore gives the theorem :— *Maximum power is transferred from a generator to the load when the resistance of the load is equal to the resistance of the generator.*

ELECTROSTATICS AND CAPACITY

In order to study the problem of capacity it is necessary first of all to consider electrostatics in a little more detail. The basis of electrostatics is "Coulomb's Law" concerning the force of repulsion between two like charges of q_1 and q_2 ESU respectively at a distance d cms. apart :—

$$F = \frac{q_1 \times q_2}{kd^2} \text{ dynes,} \tag{27}$$

where k is the "dielectric constant" of the medium separating

the charges. This constant is unity for a vacuum (and approximately so for air) and this law is the basis on which the ESU of charge was defined.

Field strength, potential difference, and potential

The "field strength" F at a point is *the force in dynes that would act on a positive unit charge placed at that point.* Field strength is measured in dynes per unit charge.

The "potential difference" between two points is *the work done (in ergs) in moving a unit positive charge from one point to the other.*

The "potential" at a point is *the work done (in ergs) in moving a unit positive charge from infinity to the point.* That is to say, it is the potential difference between the particular point and some reference point remote from the field.

Potential at a point distant r cm. from a point charge

The potential at the point B, distant r cm from a charge q ESU (*see* Fig. 89), is the work done in moving a unit positive charge from

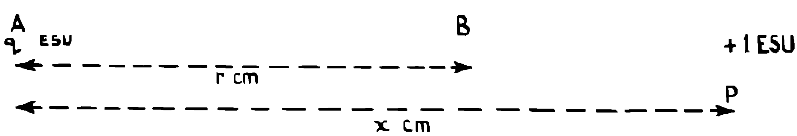


FIG. 89.—Potential at a distance r from a point charge.

infinity to the point B. Suppose that the unit charge has been moved to a point P distant x cm. from A. The field strength at P is

$$F = \frac{q \times 1}{kx^2} \text{ dynes per unit charge}$$

If the unit charge is moved a small distance δx towards B it will experience a force resisting this motion, and the force may be taken as constant and equal to $\frac{q}{kx^2}$ dynes over the *small* distance δx.

The work done in this short distance is $\frac{q}{kx^2}\,\delta x$ ergs.

The work done in bringing the unit charge from infinity to the point B, r cm. from q, is the summation of this work over the total distance, and this is equal to :—

$$\int_r^\infty \frac{q}{kx^2}\,dx, \text{ so that :—}$$

$$E = \left[-\frac{q}{kx}\right]_r^\infty$$

$$\therefore \qquad E = \frac{q}{kr} \text{ ESU} \qquad (28)$$

Since the field strength at B is :—

$$F = \frac{q}{kr^2}$$

it is seen that in this particular case :—

$$F = -\frac{dE}{dr} \tag{29}$$

i.e., Field strength = − potential gradient.

This is a completely general result, true for any electrostatic field.

Potential of an isolated conducting sphere carrying a charge

Consider a metal sphere carrying a charge of q ESU and of radius a cms. This sphere is to be imagined as completely isolated ; that is, infinitely remote from any distorting fields Under these conditions the charge will be distributed equally over the whole surface of the sphere, and it can be shown that, at an external point such as P, this distributed charge will have exactly the same effect as a charge q ESU concentrated at the centre of the sphere.

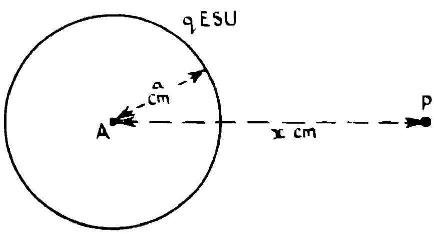


FIG. 90 Point P distant x cm from charged sphere.

The potential at the surface of the sphere is then the potential at a point distant a cms. from a point charge q ESU,

i.e., $$E = \frac{q}{ka} \text{ ESU}$$

$$\therefore \qquad \frac{q}{E} = ka \tag{30}$$

The interpretation of this result is that in the case of an isolated charged sphere the ratio of charge to potential is a constant depending only on the dimensions of the sphere and nature of the surrounding medium. This constant is called the " Capacity " of the sphere, and in a vacuum ($k = 1$) is numerically equal to the radius of the sphere in centimetres.

In general, *for any insulated conductor, the ratio of charge to potential is a constant, depending only on the shape and dimensions of the conductor and on the nature of the surrounding medium.* This ratio is called the capacity of the conductor, and denoting it by C :—

$$C = \frac{Q}{E} \tag{31}$$

Capacity has the dimensions of a length, and the electrostatic unit of capacity is the centimetre. If Q is measured in coulombs and E in volts, then the capacity C in farads is :—

$$C \text{ (farads)} = \frac{Q \text{ (coulombs)}}{E \text{ (volts)}} \tag{32}$$

Thus a conductor is said to have a capacity of one farad if a charge of one coulomb raises its potential by one volt.

$$1 \text{ farad} = 9 \times 10^{11} \text{ ESU of capacity.} \tag{33}$$

One farad is therefore the capacity of an isolated sphere of radius 9×10^{11} cms. in a vacuum, *i.e.*, $5{\cdot}6 \times 10^{6}$ miles radius. Clearly such a unit is much too large for practical purposes and the unit usually employed is the "micro-farad" (μF).

$$1 \mu\text{F} = 10^{-6} \text{ farads} = 9 \times 10^{5} \text{ ESU of capacity.} \tag{34}$$

Smaller capacities are sometimes expressed in "micro-micro-farads" ($\mu\mu$F) or pica-farads (pF).

$$1 \mu\mu\text{F} = 10^{-6}\ \mu\text{F} = 10^{-12} \text{ farads} = 0{\cdot}9 \text{ ESU of capacity.} \tag{35}$$

Capacity of a parallel plate condenser

Consider an isolated sphere of radius r cms. carrying a charge q ESU (Fig. 91). It has been seen that the potential is $\frac{q}{kr}$ ESU and the capacity is kr ESU.

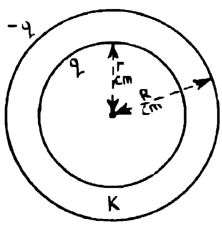


FIG. 91 Capacity between two concentric spheres.

Now suppose that the sphere is surrounded by another concentric conducting sphere of radius R cms. The potential of the outer sphere will be $\frac{q}{kR}$ ESU. If now the outer sphere is earthed it takes on an induced charge $-q$ and its potential falls to zero; that is to say, its potential is $+\frac{q}{kR}$ due to the charge on the inner sphere, and $-\frac{q}{kR}$ due to its own induced charge. But the potential inside a hollow conductor due to a charge on its surface is everywhere the same as the potential of the surface. Therefore due to the induced charge, the potential of the inner sphere is $-\frac{q}{kR}$. The effect of the two charges together is to give the outer sphere zero potential and

the inner sphere a potential $\left(\frac{q}{kr} - \frac{q}{kR}\right)$, the first term being due to the original charge, and the second term due to the induced charge on the outer sphere.

The potential difference between the spheres is :—

$$E = \frac{q}{kr} - \frac{q}{kR} = \frac{q}{k}\left(\frac{1}{r} - \frac{1}{R}\right) \text{ESU}$$

The capacity of the system is :—

$$C = \frac{q}{E} = \frac{k}{\frac{1}{r} - \frac{1}{R}} \text{ ESU}$$

$$\therefore \qquad C = \frac{kRr}{R - r} \text{ ESU} \tag{36}$$

The capacity per unit area of the inner sphere is therefore

$$\frac{kR}{(R - r)\, 4\,\pi r} \text{ ESU}$$

From this result can be deduced the capacity of a flat parallel plate condenser. For suppose that r and R are both infinitely large but that their difference $R - r = \text{d}$ is finite.

The capacity per unit area of the inner sphere becomes

$$\underset{r \to \infty}{\text{Limit}} \frac{kR}{(R - r)4\pi r} = \underset{r \to \infty}{\text{Limit}} \frac{k\,(r + \text{d})}{\text{d}4\pi r}$$

$$= \frac{k}{4\pi\text{d}}$$

This is the capacity of a parallel plate condenser per unit area. If A sq. cms. is the area of the plates,

$$C = \frac{kA}{4\pi\text{d}} \text{ ESU} \tag{37}$$

In air $\qquad k \simeq 1$

$$\therefore \qquad C_{air} \simeq \frac{A}{4\pi\text{d}}$$

This result may be obtained more easily if "Coulomb's Theorem" is assumed. This theorem is a basic theorem of electrostatics. It states that *the field strength just outside the surface of a conductor is* $F = \frac{4\pi\sigma}{k}$ *where* σ *is the charge density at that point of the surface (charge per unit area) and* k *is the dielectric constant of the surrounding medium. Moreover, the direction of the field is normal to the surface.*

Let the area of each plate be A sq. cms. It will be assumed that the area of the plates is large compared with the distance d cms. between them, so that there is everywhere between the plates a uniform normal electric field of strength F, say.

If E_p is the potential at a point such as P (Fig 92) and the field strength is F,

then
$$F = -\frac{dE_p}{dx}$$

$$\therefore \qquad E_p = -\int F\,dx$$

and the PD between the plates is —

$$E = -\int_d^0 F\,dx = Fd$$

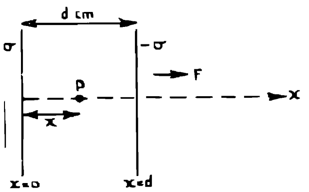


FIG 92 —Capacity between two flat parallel plates

But F is everywhere constant and by Coulomb's Theorem, near the surface, $F = \frac{4\pi\sigma}{k}$

$$\therefore \qquad E = \frac{4\pi\sigma d}{k} \quad \text{ESU}$$

But the total charge on each plate is numerically

$$q = \sigma A \quad \text{ESU}$$

$$C = \frac{q}{E} = \frac{kA}{4\pi d} \quad \text{ESU}$$

Capacity between two co-axial cylinders

To find an expression for the capacity of a condenser whose plates are co-axial cylinders, (see Fig 93), of radii r and R, (e g the capacity of a co-axial cable)

It is necessary first to consider the electrostatic field at a distance x cms from a uniformly distributed line charge of λ ESU per cm

Consider the field due to a length ds of the charge —

$$dF = \frac{\lambda\,ds}{kx^2 \sec^2\theta}$$

and the component perpendicular to the line charge is —

$$dF_x = \frac{\lambda\,ds}{kx^2 \sec^2\theta} \cos\theta$$

But
$$ds = d\,(x \,.\, \tan\theta) = x \sec^2\theta\,d\theta$$

$$\therefore \qquad dF_x = \frac{\lambda \cdot \cos\theta \cdot d\theta}{kx}$$

and the total field due to a line of infinite length having a charge λ per cm is :—

$$F = \frac{\lambda}{kx}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos\theta\, d\theta = \frac{\lambda}{kx}\left[\sin\theta\right]_{-\frac{\pi}{2}}^{+\frac{\pi}{2}}$$

$$= \frac{2\lambda}{kx} \text{ dynes per unit charge.}$$

Now consider a long cylinder of radius r cms., and carrying a charge λ ESU per cm. of its length. At points outside the cylinder,

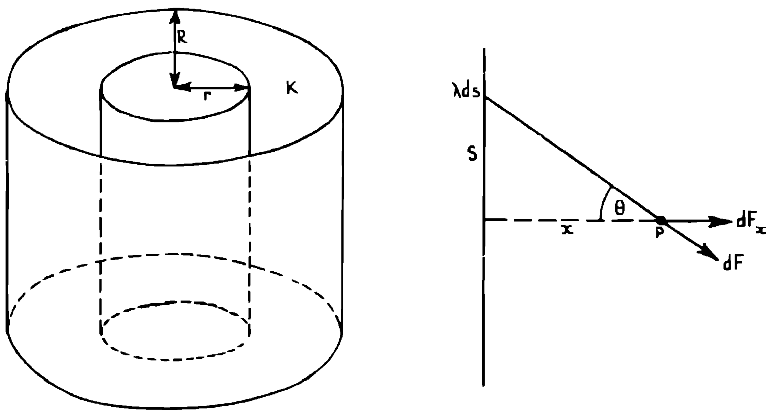


FIG. 93.—Capacity between two coaxial cylinders.

the charge will behave like a line charge λ per cm. along the axis of the cylinder, and the field at a point distant x cms. from the axis will be $\frac{2\lambda}{kx}$ ESU. The potential at such a point is therefore :-

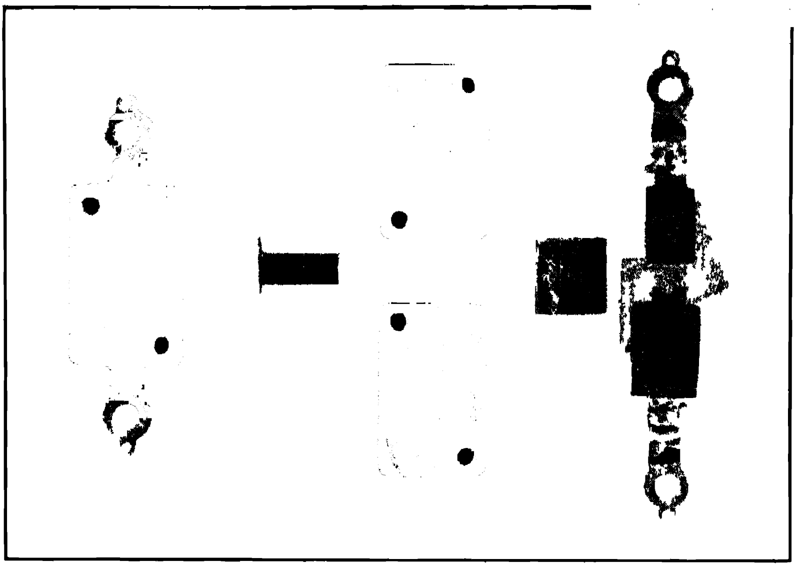
$$E = -\int F \cdot dx = -\frac{2\lambda}{k}\int \frac{1}{x}dx = -\frac{2\lambda}{k}\log_e x$$

The potential at the inner plate is therefore $-\frac{2\lambda}{k}\log_e r$,

and the potential at the outer plate is $-\frac{2\lambda}{k}\log_e R$.

The potential difference is therefore $\frac{2\lambda}{k}\log_e\left(\frac{R}{r}\right)$; but the charge per unit length is λ ESU, therefore the capacity per unit length is :—

$$C = \frac{k}{2\log_e\left(\frac{R}{r}\right)} \text{ ESU per cm.} \qquad (38)$$



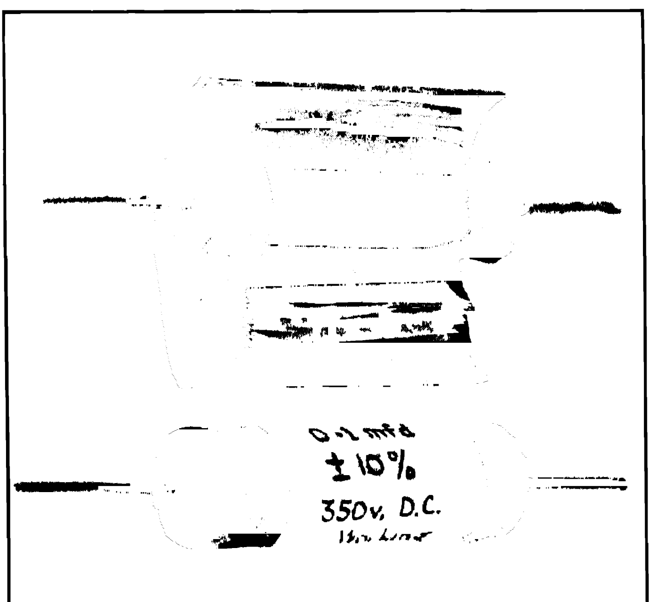


PLATE 3.—Mica and paper condensers showing construction

Capacity between two parallel wires

If two wires both have radius a, and their centres are separated by a distance c (*see* Fig. 94), it can be shown that the capacity

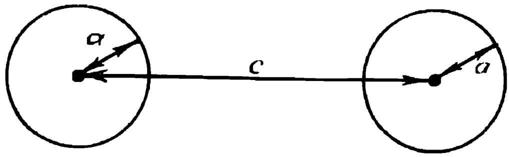


FIG. 94.—Capacity between two parallel wires.

between the wires is given by :—

$$C = \frac{k}{4 \log_e \left(\frac{c}{a}\right)} \text{ ESU per cm.}$$

$$= \frac{0 \cdot 01941k}{\log_{10}\left(\frac{c}{a}\right)} \ \mu\text{F per mile} \qquad (39)$$

provided that $a \ll c$.

Dielectric constant

The " dielectric constant " k of a material may be defined as *the ratio of the capacity of a condenser employing that material as dielectric to the capacity of an exactly similar condenser but employing a vacuum as the dielectric.*

Some typical values of dielectric constants are given in Table VI:—

TABLE VI.

Dielectric constants.

Medium	Dielectric constant
Paxolin	5 to 8
Ebonite	2·7 to 2·9
Mica .. .	5·7 to 7
Polythene ..	2·2 to 2·4
Paraffin wax .	2 to 2·3
Air	1·0006
Vacuum .. .	1

From the formula for the capacity of a parallel plate condenser it will be seen that to give a large value of capacity the conditions are : large area plates, small separation between plates, and a high dielectric constant.

Dielectrics

It has already been mentioned that when an electrostatic field is applied to a dielectric there is a distortion of the electron orbits due to the interaction of the field and the orbital electrons. This results in a mechanical stress being set up in the dielectric, and a dielectric in this condition is said to be "polarised". This hypothesis explains all the phenomena associated with dielectrics.

If a condenser is charged and then discharged, and is then left for a short time with the plates open-circuited, it is found that it again becomes charged to a small extent. This charge, which is of the same polarity as the original charge, is called the "residual charge", and is due to the slow return of the polarised dielectric to its normal unpolarised condition.

When a dielectric is polarised, there is a nett transfer of electrons in the direction opposite to that of the applied field. This transfer of electrons gives rise to a "displacement current", and the mechanical stress produces heat in the dielectric. This heat produced indicates a power loss known as "dielectric loss".

If a condenser is subjected to a very high voltage, "dielectric breakdown" may occur. This is due to the large field overcoming the forces holding the electrons in their orbits; the electrons break away, and the dielectric becomes a conductor. With solid dielectrics the damage caused is permanent, and the condenser becomes useless. In the case of liquid and gaseous dielectrics, sparking occurs, but as soon as the peak voltage has passed, the dielectric "heals up" and recovers its normal properties. The ability of a material to resist breakdown is known as its "dielectric strength", and is measured in terms of the voltage at which breakdown occurs. Such figures are useful only for rough comparison, since they depend on the thickness of the sample and on the conditions under which the test is made. Considerations of dielectric strength are mainly the concern of the power engineer.

The type of condenser used for any particular purpose depends largely on two factors: the capacity required, and the maximum voltage to which the condenser will be subjected. Air condensers can be used for capacities up to 0·001 μF (1,000μμF), but they are bulky and are used only where a variable condenser is required. Ceramic condensers cover a similar range, 5μμF to 1,000μμF, for fixed condensers. Mica condensers are normally used over a range from 50μμF to 0·01μF. Paper condensers cover the range from 0·0001μF to 8μF. Where possible, paper condensers are used for economic reasons; but since the dielectric strength of mica is approximately ten times that of paper, a paper condenser of say 0·01μF for 1,000 V. working might be larger and more expensive than a similarly rated mica condenser.

Electrolytic condensers

If two aluminium plates are immersed in a suitable electrolyte

such as a borax solution, the application of a direct voltage will cause a current to flow. This current rapidly diminishes to a small value due to the formation of a thin insulating film of aluminium oxide on the positive electrode. Owing to the extreme thinness of this film ($0 \cdot 01$ to $0 \cdot 05 \times 10^{-3}$ cm.), a high capacity will exist between the positive electrode and the solution. This arrangement forms the basis of electrolytic condensers enabling

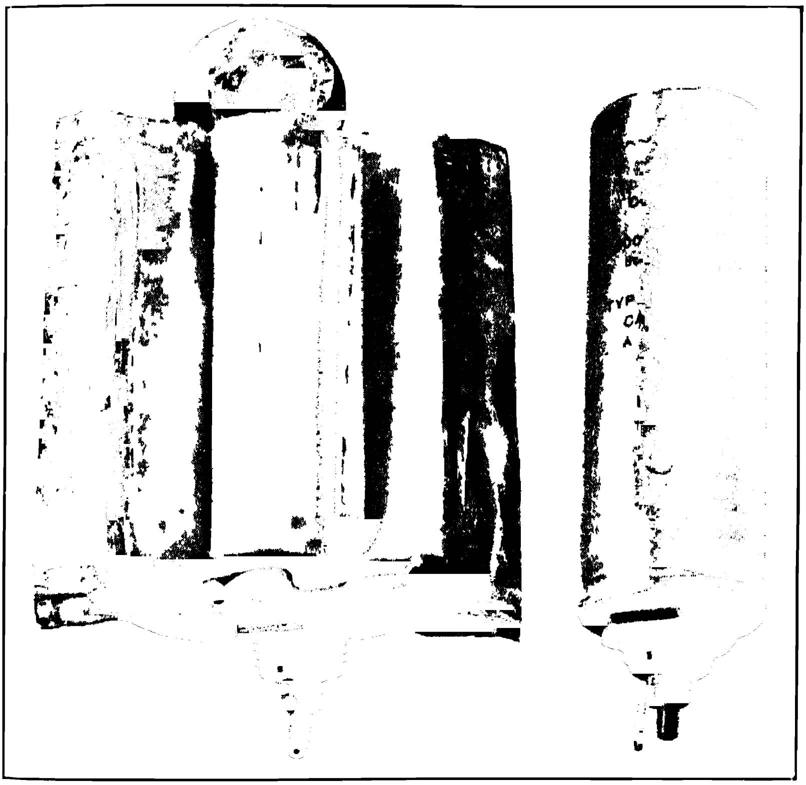


PLATE 4.—Electrolytic condenser.

large values of capacity to be obtained in a small volume and at low cost.

Most electrolytic condensers require the presence of a polarising direct current to maintain the film, and such condensers can be used only in circuits where the peak value of the alternating voltage is less than that of the superimposed direct voltage.

The dielectric strength of the oxide film is high, and condensers can be made to withstand voltages up to 500–700 V. The rated voltage of a condenser is always slightly less than that of the forming

voltage. Above this voltage, the leakage current becomes excessive, and perforations of the film may occur.

A film of electrolyte is always present, so that any perforation that may occur can be immediately resealed. The nature of the electrolyte may vary, giving three different types :—

(*a*) The " wet " electrolytic condenser. In this type, the anode is suspended in the centre of a cylindrical container which forms the cathode, and a liquid electrolyte is used. The shape of the anode is designed to give a large surface area, thus increasing the capacity, while the liquid electrolyte gives the advantage of quick resealing after a breakdown.

(*b*) The " semi-dry " electrolytic condenser. The electrodes take the form of long strips of aluminium foil, the dielectric film having been formed on the anode. These are rolled together, being separated usually by a cotton gauze impregnated with the electrolyte. The roll is mounted in an aluminium or bakelite container.

(*c*) The " dry " electrolytic condenser. The construction is similar to that of the semi-dry type, but a material is added to the electrolyte to make it solid at normal temperatures. This is the most common type used in this country.

Reversible electrolytic condensers consist of ordinary electrolytic condensers with an oxide film formed on both electrodes. They are thus equivalent to two condensers in series, and have half the capacity of an ordinary condenser of the same size.

Condensers in parallel

When condensers are connected in parallel, they have a common voltage, but each has its own charge.

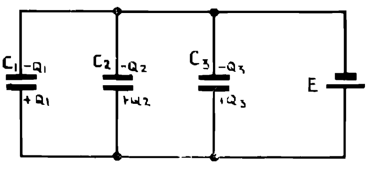


FIG. 95 – Condensers in parallel.

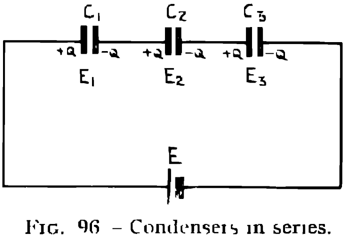


FIG. 96 – Condensers in series.

The total charge $\quad Q = Q_1 + Q_2 + Q_3$

$$\therefore \quad \frac{Q}{E} = \frac{Q_1}{E} + \frac{Q_2}{E} + \frac{Q_3}{E}$$

$$\therefore \quad C = C_1 + C_2 + C_3 \qquad (40)$$

Condensers in series

A charge $+Q$ on the left-hand plate of C_1 will induce $-Q$ on the second plate, this causes $+Q$ on C_2, and so on. Thus all the

condensers have the same *charge*, which is equal to the total charge :—

$$Q = Q_1 = Q_2 = Q_3$$

But $$E = E_1 + E_2 + E_3$$

$\therefore$ $$\frac{Q}{C} = \frac{Q_1}{C_1} + \frac{Q_2}{C_2} + \frac{Q_3}{C_3}$$

$\therefore$ $$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} \qquad (41)$$

Charge and discharge of a condenser

Consider a series circuit of battery, condenser, resistance and key as shown in Fig. 97. The left-hand plate of the condenser will

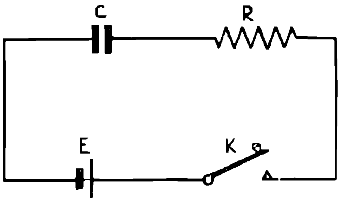


FIG. 97 Circuit for charging a condenser C through a resistance R.

be at the same potential as the negative plate of the battery, and on closing the key the whole of the EMF will drive a current through R to charge the condenser. The magnitude of this current is dependent on R.

Immediately C commences to acquire a charge, the charging current will drop, and therefore the PD across R will fall. This process continues so that, as the charge in the condenser approaches its full value, the charging current becomes less and less. By continuation of this argument it may be shown that, in theory, the condenser will take an infinite time to become fully charged. This deduction is proved mathematically below.

At any instant, t seconds after closing the key,

let v be the PD across the condenser,

i be the charging current,

q be the charge on the condenser.

Then $$i = \frac{dq}{dt}$$

$$= \frac{d\,(Cv)}{dt}$$

But $$i = \frac{\text{PD across } R}{R}$$

$$= \frac{E - v}{R}$$

$$\therefore \qquad \frac{E-v}{R}=C\frac{dv}{dt}$$

$$\frac{dv}{dt}+v\cdot\frac{1}{CR}=\frac{E}{CR}$$

Multiply by $e^{\frac{t}{CR}}$:—

$$\frac{dv}{dt}\cdot e^{\frac{t}{CR}}+\frac{v}{CR}\cdot e^{\frac{t}{CR}}=\frac{E}{CR}\cdot e^{\frac{t}{CR}}$$

The left-hand side of this equation is the differential coefficient of $v.\,e^{\frac{t}{CR}}$, so that this equation can be written as :—

$$\frac{d}{dt}\left[v\cdot e^{\frac{t}{CR}}\right]=\frac{E}{CR}\cdot e^{\frac{t}{CR}}$$

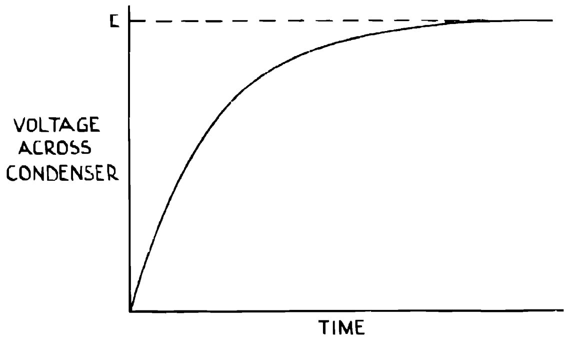


FIG. 98.—Graph showing charging of condenser.

Integrating :—

$$v\cdot e^{\frac{t}{CR}}=E\cdot e^{\frac{t}{CR}}+K, \text{ where } K \text{ is a constant,}$$

$$\therefore \qquad v=E+K\cdot e^{-\frac{t}{CR}}$$

But when $t=0$, $v=0$, hence K may be determined :—

$$K=-E$$

Thus
$$v=E\left(1-e^{-\frac{t}{CR}}\right) \qquad (42)$$

From this formula it will be seen that, as t increases, v becomes nearer and nearer to E but will not reach it until, t being infinite, the factor $e^{-\frac{t}{CR}}$ becomes zero. A graph showing the voltage across the condenser increasing exponentially with time is given in Fig. 98

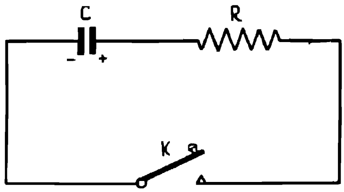


FIG. 99.—Circuit for discharging a condenser C through a resistance R

If the battery is now removed and the key again closed, the condenser starts to discharge. At first, the PD across the condenser driving the discharge current through R is equal to E; but as soon as the condenser partly discharges, this PD drops, and the current (*i.e.* the rate of discharge) drops. The curve of the discharge is again exponential, and theoretically the condenser never fully discharges.

The current i is now decreasing, so that :—

$$i = -\frac{dq}{dt}$$

$$= -C\frac{dv}{dt}$$

Also $$i = \frac{\text{PD across } R}{R}$$

$$= \frac{v}{R}$$

$\therefore$ $$\frac{v}{R} = -C\frac{dv}{dt}$$

$\therefore$ $$\frac{dv}{dt} + \frac{1}{CR} \cdot v = 0$$

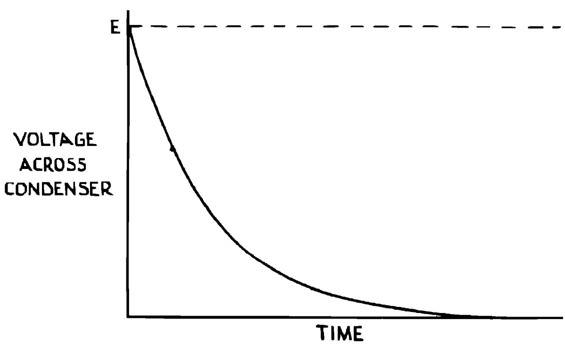


FIG. 100.—Graph showing discharge of condenser.

$$\therefore \qquad \frac{dv}{dt}e^{\frac{t}{CR}} + \frac{1}{CR}v \, . \, e^{\frac{t}{CR}} = 0$$

$$\therefore \qquad \frac{d}{dt}\left(v \, . \, e^{\frac{t}{CR}}\right) = 0$$

$$\therefore \qquad v \, . \, e^{\frac{t}{CR}} = K, \text{ where } K \text{ is a constant}$$

$$\therefore \qquad v = K \, . \, e^{-\frac{t}{CR}}$$

When $t = 0$, $v = E$ hence K may be determined —

$$K = E$$

Thus
$$v = E \, . \, e^{-\frac{t}{CR}} \qquad (43)$$

The graph of this function gives therefore the discharge curve of the condenser, and is shown in Fig 100

Time constant

In the design of electrical apparatus it is often required to produce a resistance capacity circuit with a definite time of discharge or charge. It has been shown that theoretically this time is infinite,

TABLE VII

Condenser charge and discharge

% diverging from total charge or discharge	Time of Charge or Discharge
1.8%	4 CR secs
3.1%	3.5 CR secs
5.0%	3 CR secs
8.3%	2.5 CR secs
13.5%	2 CR secs
37 %	1 CR secs

but the practical aspect will now be considered From equations 42 and 43 for v, it is noticed that the rate of charge or discharge is dependent on the product CR This product is therefore termed the "time constant" of the circuit

(*a*) *Charge* · When $t = CR$ secs

$$v = E\,(1 - e^{-1})$$
$$= 63\% \; E.$$

(*b*) *Discharge*. When $t = CR$ secs.

$$v = E \, . \, e^{-1}$$
$$= 37\% \; E$$

That is, when C and R are in farads and ohms respectively, the product CR gives the time in seconds for the condenser (*a*) to charge up to 63 per cent. of full charge, or (*b*) to discharge from full charge to 37 per cent. of full charge. Table VII shows how this time constant is utilised in designing practical circuits when a certain margin is permissible between the total charge or discharge and that actually reached after a given time.

Example.—A condenser-resistance circuit is required to have discharge time of 150 milliseconds. A 5 per cent. margin is allowable. (Condenser must be at least 95 per cent. discharged in the 150 milliseconds.)

From table, for a 5 per cent. margin,

$$3\,CR = 0 \cdot 150 \text{ secs.}$$

$$\therefore \qquad CR = 0 \cdot 050 \text{ secs}$$

Choosing $C = 0 \cdot 1 \mu F$ as a suitable condenser,

$$R = \frac{0 \cdot 050}{0 \cdot 1 \times 10^{-6}} \text{ ohms} = 0 \cdot 5 \text{ Megohm. } Ans.$$

MAGNETISM

Certain specimens of magnetite, an ore of iron mined in various parts of the world, are called natural magnets or lodestones, and possess the following properties:—

(*a*) they attract small fragments of iron and steel;

(*b*) when suitably suspended they come to rest in a definite position relative to the points of the compass;

(*c*) they are able to confer both these properties on certain other materials, notably iron, steel, nickel and cobalt.

These properties have been known from the earliest times, but it was not until very much later that it was found that these properties could be artificially imparted to steel and iron by means of an electric current.

Magnets may be classified as permanent magnets and electromagnets. Permanent magnets are made of steel or such alloys as cobalt steel, and once magnetised they retain their magnetic properties for a long period under normal conditions. Electromagnets are made with a core of soft iron or of iron alloys such as permalloy (iron-nickel), and have the property that, although more easily magnetised, they lose their magnetic properties almost immediately when the magnetising influence is removed.

Permanent magnets

Fig. 101 represents a rectangular bar magnet that will attract fragments of iron brought near either end, and will exert a force of either repulsion or attraction on other magnets in the vicinity. The influence of the magnet may be detected in the surrounding space in various ways, and it is found to vary inversely with the square of the distance from the magnet. To account for this

phenomenon, the magnet is said to establish a magnetic field, which is represented by the curved lines (lines of force) in Fig. 101. This method of representing the magnetic field is merely a convention adopted to give a simple pictorial representation of the influence of the magnet in the surrounding space.

Ewing has shown that the behaviour of permanent magnets and magnetic materials may be explained on the assumption that they are made up of a very large number of small cuboids, each of which has the properties of a permanent magnet. In the unmagnetised material, these small magnets have a random orientation, and the specimen shows no resultant magnetic properties. Magnetisation has the effect of orientating the small magnets along an axis, called the magnetic axis, which joins those two points near the ends of the

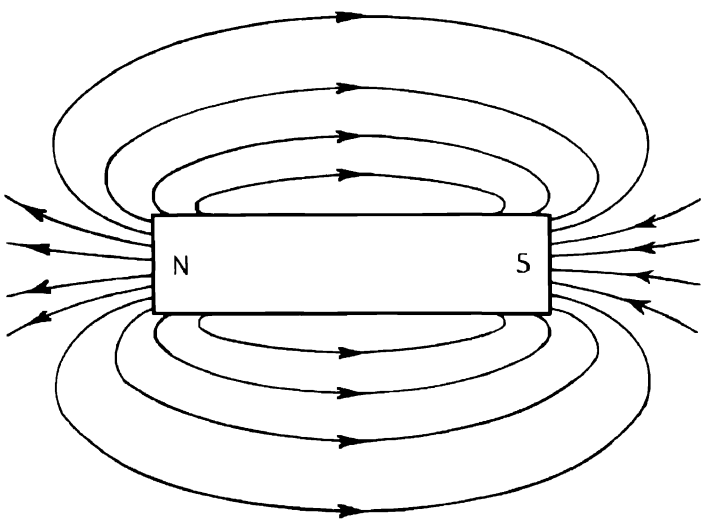


FIG. 101.—Field round a bar magnet.

magnet where the effect of the magnetic field appears to be concentrated; these points are known as poles. Every magnet is assumed to possess two poles, a north-seeking pole, which experiences a force of attraction towards the Earth's north magnetic pole, and a south-seeking pole, which is attracted to the Earth's south magnetic pole. These poles cannot, of course, be isolated, since two unlike poles comprise a magnet and an isolated pole has no physical meaning. It is possible, however, to visualise an isolated pole by considering a magnet that is so long, that the influence of the unwanted pole is negligible in the vicinity of the pole being examined. On this basis, careful experiment has shown that the force between two such poles is directly proportional to the "strength" m of each pole, and inversely proportional to the square of the distance d between the poles (*see* Fig. 102),

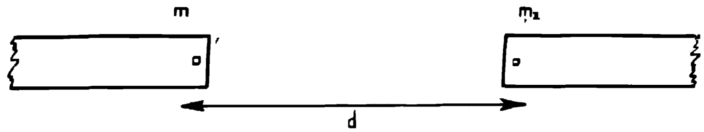


FIG. 102.—Force between two magnetic poles distant d apart.

i.e., $$\text{Force} \propto \frac{mm_1}{d^2}$$

In this connection, the strength of a magnetic pole is merely a theoretical idea, but it can be given a physical meaning by defining a unit magnetic pole as follows.

A " unit magnetic pole " is *that pole which, when separated by one centimetre (in vacuo) from an exactly similar pole, repels it with a force of one dyne.*

The law for the force of repulsion then becomes :—

$$\text{Force} = \frac{mm_1}{\mu d^2} \text{ dynes} \tag{44}$$

where d is the distance between the poles in centimetres, and μ is a property of the surrounding medium, known as its permeability ($\mu = 1$ for a vacuum, and is approximately equal to 1 for air). For convenience, north-seeking poles are assumed to have a positive pole strength, and south-seeking poles a negative pole strength. Thus, for unlike poles, the force is a negative repulsion, *i.e.*, an attraction.

The " field strength " at a point is *the force that would be experienced by a unit north-seeking pole placed at that point.* It is a vector, denoted by H, and its magnitude is measured in dynes per unit pole, gauss, or occasionally oersteds.

The force experienced by a pole of strength m will be Hm dynes. A magnet of pole strength m and length l cms. set at right angles

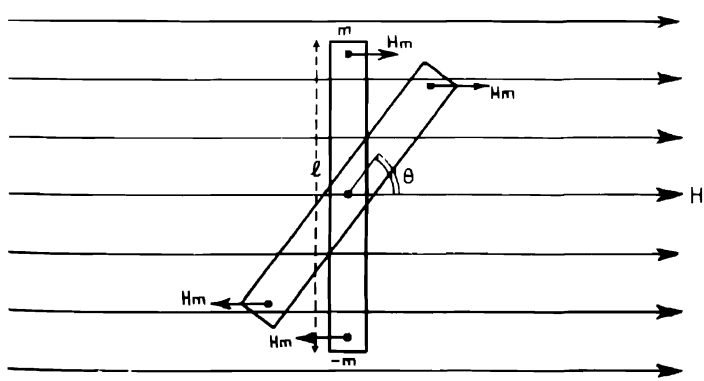


FIG. 103.—Magnetic moment.

to a uniform magnetic field of strength H gauss will therefore experience a turning moment Hml centimetre-dynes tending to align it with the magnetic field. When the magnet makes an angle θ with the field (*see* Fig. 103), the turning moment is reduced to $Hml \sin \theta$, which vanishes when $\theta = 0$.

The product ml is known as the "magnetic moment" of the magnet, and is *the turning moment that is experienced by the magnet when placed at right angles to a uniform magnetic field of unit strength.*

Magnetic flux and lines of induction

From equation 44, the force between two poles, one a unit pole, the other of strength m, at a distance d cms. apart, is :—

$$\text{Force} = \frac{m}{\mu d^2} \text{ dynes.}$$

This, by definition, is the field strength at a point d cms. from an isolated pole of strength m,

i.e. $$H = \frac{m}{\mu d^2} \text{ gauss.}$$

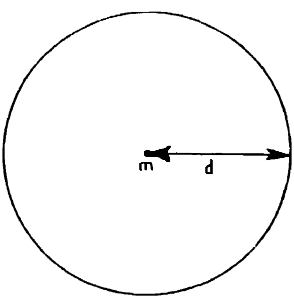


FIG. 104.—Sphere of radius d surrounding isolated pole m.

Thus, the field strength in this simple case depends on the permeability of the medium, and this is true in the general case, for a complex magnetic field may be built up of fields due to simple poles. It is convenient to have a notation for expressing magnetic effects that is independent of the permeability of the medium. This is done by using a theoretical concept, lines of induction, defined as follows :—

4π lines of induction leave every unit north-seeking pole and enter every unit south-seeking pole.

The total number of lines of induction in a magnetic system is called the "flux", represented by the letter Φ. The flux density B is the flux per unit area, and is expressed in lines per square centimetre. Thus $\Phi = B.A$, where A is the cross-sectional area normal to the lines of induction.

The fact that B, like Φ, is a vector may be seen from the

fact that it is the number of lines per square cm. crossing a surface that is orientated in a particular way; it may therefore be regarded as having direction as well as magnitude.

Consider an imaginary spherical surface of radius d cms. having an isolated pole of strength m at its centre (Fig. 104). Then since by definition $4\pi m$ lines of induction leave this pole, and can only terminate in an unlike pole (assumed very remote), all these lines must cross the imaginary surface. Since the pole is isolated, it may be assumed that the lines will be symmetrically distributed, giving a flux density of :—

$$B = \frac{4\pi m}{4\pi d^2} \text{ lines per square cm., } i.e., B = \frac{m}{d^2} = \mu \frac{m}{\mu d^2} = \mu H.$$

Thus in the particular case of an isolated pole, $B = \mu H$. This is true in the general case, since a complex field may be regarded as being built up of simple poles.

Hence flux, flux density, and field strength are connected by the relationship :—

$$\Phi = A.B = A\mu H \tag{45}$$

Note that in a vacuum ($\mu = 1$) *B and H are vectors of equal magnitude; they have always the same direction, and the lines of induction and lines of force may be considered to be identical. This is approximately true for air.*

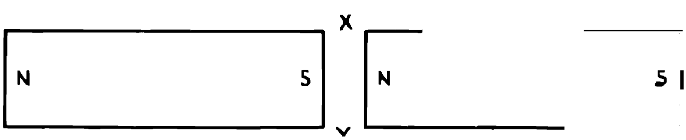


FIG. 105.—Bar magnet cut in half along XY.

Suppose that a bar magnet of pole strength m is divided by a cut XY (*see* Fig. 105) at right angles to its magnetic axis. The result will be two bar magnets of pole strength m. That is to say, the left-hand face of the gap will be a south pole and the right-hand face a north pole. The "intensity of magnetisation" I of the magnet in the region of this gap is defined as *the pole strength of either face divided by the area of the face* I is a vector, since it is associated with direction, the imaginary cut having been made at right angles to the magnetic axis. The magnitude of I is measured in unit poles per square cm.

Behaviour of soft iron in a magnetic field

Consider a bar of soft iron of permeability μ placed in a uniform magnetic field of strength H gauss; for simplicity, the bar is placed with its length in the direction of the field. It is observed in practice that the soft iron becomes magnetised, with north and south poles as shown in Fig. 106, the axis of the resultant magnet being in the direction of the magnetising field.

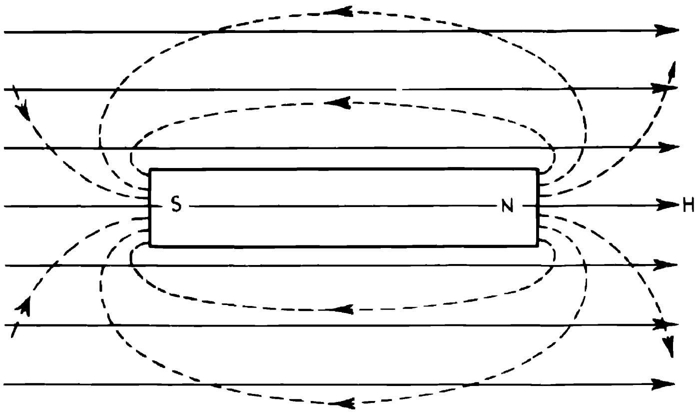


Fig. 106.—Soft iron bar in magnetic field.

The field in the vicinity of the soft iron may be predicted by considering lines of induction. The lines of induction due to the uniform field H may be represented by straight parallel lines, those due to the induced magnetism of the soft iron will be as shown dotted in Fig. 106.

The resultant lines of induction are as shown in Fig. 107, and may be found by considering the two fields of Fig. 106 superposed. Clearly at the ends of the magnet the induction due to the two component fields is additive, whilst at the sides they are in opposition. The soft iron is therefore seen to have the effect of concentrating the lines of induction so that they pass through the soft iron in preference to passing through the air.

Now suppose that a minute gap XY is cut at right angles to the magnetic axis and that the intensity of magnetisation in this region is I. Let the pole strength of the faces of the gap be m and the

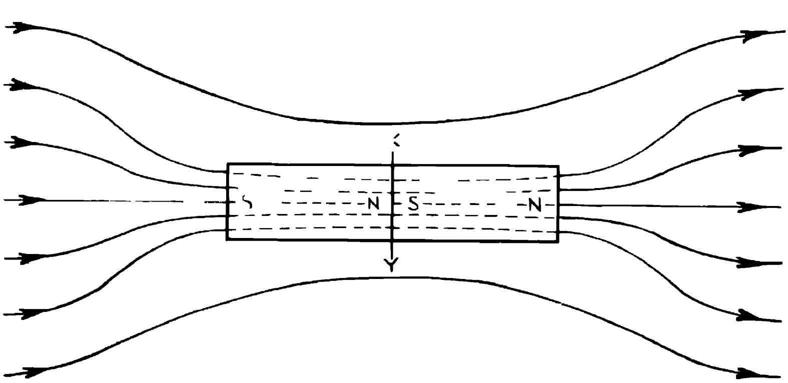


Fig. 107.—Magnetic field around soft iron bar.

area a sq cms Then by definition, $I = \frac{m}{a}$, but 4π lines of induction leave every unit north pole on the left side of the gap and enter every unit south pole on the right-hand side The total flux in the gap due to the induced magnetism of the soft iron is $4\pi m$ and the flux density due to the induced magnetism is $\frac{4\pi m}{a} = 4\pi I$

The total flux density in the gap however contains a component due to the original magnetising field H and numerically equal to H Thus the total flux density is given by

$$B = H + 4\pi I \tag{46}$$

This result has been obtained in the special case of a bar of soft iron in a uniform magnetic field It may, however, be extended to any shaped piece of iron in any type of field since the iron may be regarded as made up of a large number of very small bars, for each of which the field may be regarded as uniform

Dividing equation 46 by H —

$$\frac{B}{H} = 1 + 4\pi \frac{I}{H}$$

i e

$$\mu = 1 + 4\pi k \tag{47}$$

where k is defined as the "susceptibility" of the iron and gives a measure of the ease with which it may be magnetised by induction Equation 47 shows that susceptibility and permeability are related

Hysteresis

Fig 108 shows the cycle followed by the magnetic flux (B) produced by a bar as the magnetising field (H) is varied Starting

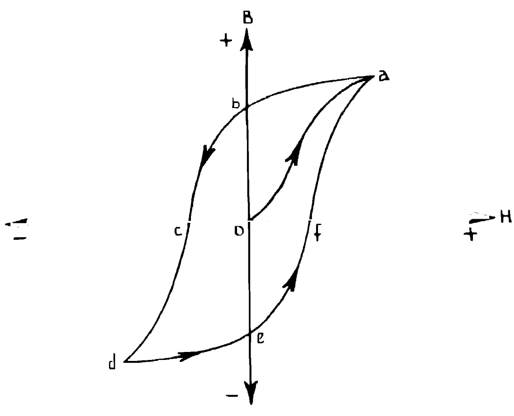


Fig 108 —Hysteresis cycle

with zero field and the iron demagnetised (point O), the field is gradually increased in the positive direction At first the magnetisation is slow, then increases rapidly until, having almost reached

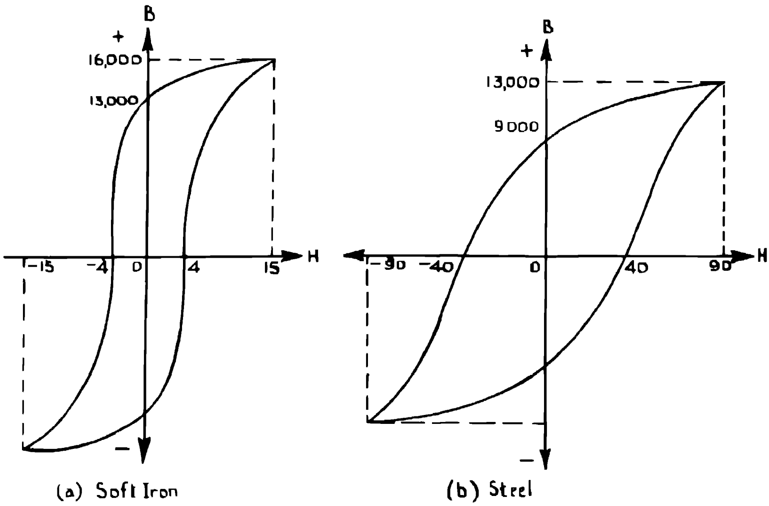


FIG 109.—B–H or hysteresis curves for soft iron and steel.

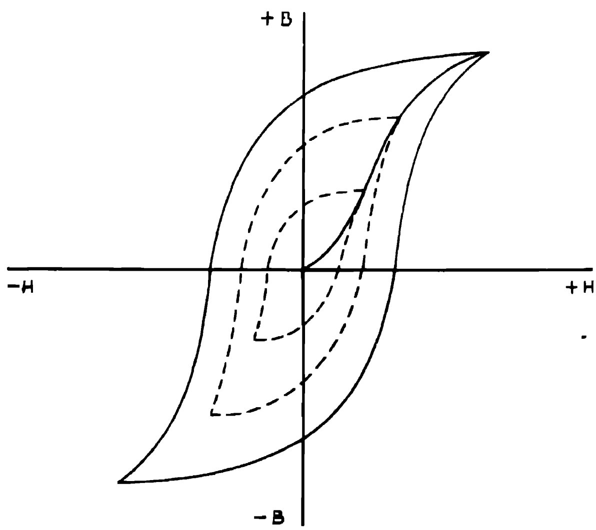


FIG. 110.—Effect of degree of magnetisation on hysteresis loop.

saturation point, the rate of change of flux drops.

The magnetising field is now reduced; the flux, however, does not follow the original curve, but decreases far more slowly until, when the field is fully removed, the iron has an amount of residual magnetism represented by Ob. Reversing the magnetising field eventually removes all the magnetism (point c) and then remagnetises the iron in the reverse direction, until another saturation point is reached at d. Returning the field to zero and then increasing it in the original direction will complete the cycle. The outer curve (a b c d e f a) so traced will be followed for all future magnetising cycles.

It will be observed from the diagram that, after the initial magnetisation, the flux may be considered as lagging behind the field; this phenomenon is termed " hysteresis ".

The distance Ob on the curve is a measure of the " retentivity " of the iron, while the distance Oc, *i.e.*, the field required to demagnetise, is a measure of the " coercivity ". These two qualities vary for different magnetic materials; curves for soft iron and steel are shown in Fig. 109.

Fig. 110 shows hysteresis curves for a specimen using different degrees of magnetisation.

Hysteresis power loss

The hysteresis loop has a further quantitative interest since it can be shown, as below, that the energy expended in performing one cycle of magnetisation is represented by the area enclosed by the curve. Thus it can be seen that more work is done in the case of steel than in the case of soft iron. This is what would be expected from the known qualities of these two metals.

This work is of importance in such apparatus as the transformer, where repeated cycles are made, for the power loss caused by hysteresis is wasted as heat.

The most convenient method of producing the reversals of magnetic field is the use of the solenoid. The hysteresis loss in a sample of magnetic material placed at the centre of a solenoid will therefore be considered.

The field (H) at the centre of the solenoid is given by the formula:—

$$H = 4\pi Ni = \frac{4\pi Ti}{l} \quad \text{(This result is verified on page 163.)}$$

Where T is the total number of turns,
N is the number of turns per unit length,
i is the current in the solenoid (in EMU),
and l is the length (in cm.).

$$\therefore \qquad i = \frac{l}{4\pi T} \times H$$

The total flux Φ through the solenoid is $B \cdot A$, where A is the cross-sectional area, and the EMF induced (*see* p 167) is :—

$$e = -T \times \frac{d(BA)}{dt} \quad \text{(in EMU)}$$

$$\therefore \qquad e = -TA \times \frac{dB}{dt}$$

The work done to complete the hysteresis cycle is done against this EMF,

$$\therefore \text{ instantaneous power } = -e \times i$$

$$= \left(TA \cdot \frac{dB}{dt}\right) \times \left(\frac{l}{4\pi T} \cdot H\right)$$

$$= \frac{Al}{4\pi} \times H \frac{dB}{dt}$$

The work dW done in a small time dt is :—

$$dW = \frac{Al}{4\pi} \times H \frac{dB}{dt} \cdot dt$$

Total work done W is $W = \dfrac{Al}{4\pi} \times \displaystyle\int H\, dB$ (48)

and it will be noticed that the integral included in the last equation is the area of the hysteresis loop This is illustrated in Fig. 111.

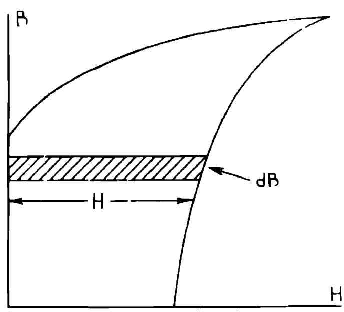


FIG 111 —Calculation of power loss due to hysteresis.

Furthermore, Al is the volume of the sample,

$$\therefore \text{ the work done per c c. } = \frac{\text{Area of loop}}{4\pi} \qquad (49)$$

If B and H are in the units of lines per sq. cm , this expression will give the work in ergs per c.c per cycle.

Now, as already stated, this work done against hysteresis is, in practice, a loss of power and therefore of much importance. If the shape of the loop is known, then the area will give a value for this loss. This method is not suitable for use on machines and an

approximate method devised by Steinmetz is preferable. Steinmetz discovered that the hysteresis loss P_H per c.c. per cycle was very closely given by the formula $P_H = \eta B_{max}^{1 \cdot 6}$ ergs, where η is a constant called the "hysteresis coefficient" and whose value depends upon the material, and B_{max} is the maximum flux density during the cycle. Hence the loss in a core of volume V c.c. over f cycles is given by :—

$$\text{Work done} = \eta \, . \, B_{max}^{1 \cdot 6} \, . \, V \, . \, f \text{ ergs}$$
$$= \eta \, . \, V \, . \, f \, . \, B_{max}^{1 \cdot 6} \times 10^{-7} \text{ joules.}$$

If f cycles per second is the frequency, then this loss occurs each second, and :—

$$\text{Hysteresis power loss} = \eta \, . \, V \, . \, f \, . \, B_{max}^{1 \cdot 6} \times 10^{-7} \text{ watts} \tag{50}$$

Dia- and para-magnetism

The type of magnetism already dealt with, which is the most important, is called *ferromagnetism*. There are, however, two other classifications which deserve brief mention here. They are *diamagnetism* and *paramagnetism*.

If a material has a permeability of less than 1, it is said to be *diamagnetic* ; if the permeability is greater than 1, the material is said to be *paramagnetic*. Examples of each are shown in Table VIII.

TABLE VIII

Diamagnetic and paramagnetic substances

Diamagnetic substances	Paramagnetic substances
Bismuth	Liquid Oxygen
Water	Air
Quartz	Palladium
Lead	Platinum
Copper	Aluminium
Hydrogen	Oxygen

The permeability of all these substances is very near unity, however ; for example, μ for platinum is $1 \cdot 000017$, and for bismuth $0 \cdot 99996$. To the paramagnetic substances may be added a class of certain ions that show far stronger paramagnetism when in the form of salts or in solution. Thus some copper salts have values of μ around $1 \cdot 75$.

The ferromagnetic substances (nickel, cobalt and iron) have far higher permeabilities, from 250 up to many thousands, and may be considered as extreme cases of paramagnetism.

Fig. 112 shows $B-H$ and $\mu-H$ curves for a typical ferromagnetic material. It will be seen that μ is by no means constant.

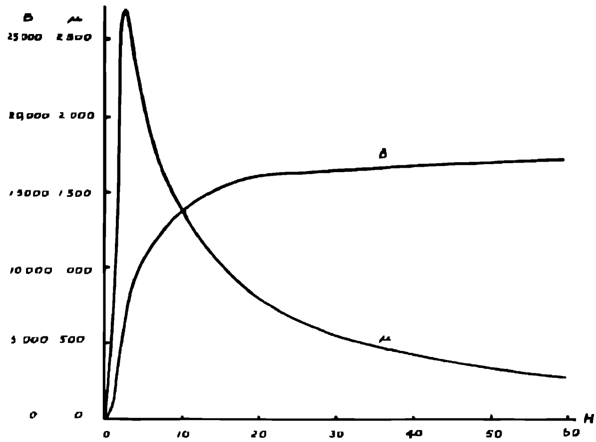


FIG 112.—Typical μ-H and B-H curves

ELECTROMAGNETISM AND INDUCTANCE

The magnetic field due to a current in a straight wire is circular around the wire, and its direction is clockwise when viewed in the direction in which the current is flowing The field strength at P

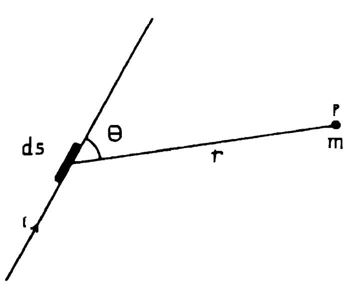


FIG 113.—Field distant r from a wire carrying current.

due to a short section of wire ds carrying a current i (Fig. 113) is given by Laplace's Law —

$$H = \frac{i \, . \, ds \, . \, \sin \theta}{r^2} \text{ gauss} \qquad (51)$$

Where i is the current,
ds is the length of the segment of wire,
θ is the angle between segment of wire and line to P,
r is the distance from segment to P.

Field at centre of coil

By Laplace's Law, the force on a unit pole at the centre of a coil in vacuo due to current i flowing through a small segment ds of that coil (*see* Fig. 114) is given by :—

$$f = \frac{ids}{r^2} \text{ dynes}$$

∴ The total force on the unit pole due to the whole coil is :—

$$F = \Sigma f$$

$$= \int_0^{2\pi r} \frac{ids}{r^2}$$

$$= \frac{2\pi i}{r} \text{ dynes}$$

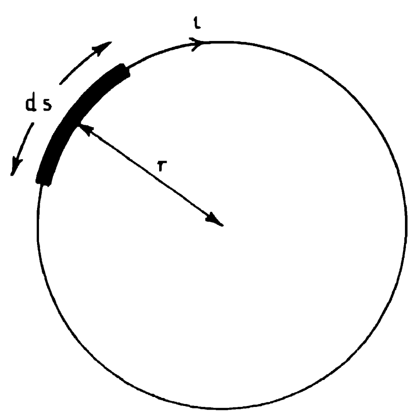


FIG. 114.—Field at centre of a coil.

Thus, if i is the current in EMU (defined on page 131), the field strength is :—

$$H = \frac{2\pi i}{r} \text{gauss}$$

Since 1 EMU of current = 10 amps,

$$H = \frac{2\pi I}{10r} \text{ gauss} \tag{52}$$

where I is in amperes.

For a coil of T turns,

$$H = \frac{2\pi I T}{10r} \text{ gauss} \tag{53}$$

The flux at the centre of the coil is therefore given by :—

$$\Phi = A \mu H = \frac{2\pi A\mu}{10 r} . T.I \tag{54}$$

showing that the flux is directly proportional to the current.

Field at centre of solenoid

Consider the case of a solenoid. Referring to Fig. 115, P is a point on the axis of a long solenoid, and AB is one turn of that solenoid.

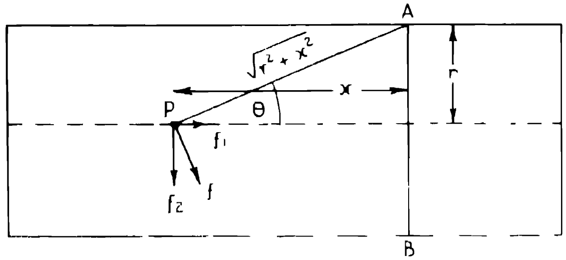


FIG. 115.—Force on unit pole at centre of solenoid, due to one turn

Then the force f on a unit pole at P due to a segment ds of AB is :—

$$f = \frac{i\,ds}{r^2 + x^2}$$

But this force can be resolved into two components, f_1 and f_2, in the directions shown, so that :—

$$f_1 = \frac{i\,.\,ds}{r^2 + x^2} \sin\theta$$

$$f_2 = \frac{i\,.\,ds}{r^2 + x^2} \cos\theta$$

To find the field at P due to the turn AB, the forces f must be summed over the complete turn. The components f_2 will clearly have a zero resultant, but :—

$$\Sigma f_1 = \Sigma \frac{i\,.\,ds}{r^2 + x^2} \sin\theta$$

$$= \frac{2\pi r\,.\,i\,.\,\sin\theta}{r^2 + x^2} \text{ dynes}$$

But $$\sin\theta = \frac{r}{\sqrt{r^2 + x^2}}$$

hence $$\Sigma f_1 = \frac{2\pi r\,.\,i\,.\,\sin\theta}{r^2 + x^2} = \frac{2\pi r^2\,i}{(r^2 + x^2)^{\frac{3}{2}}} \text{ dynes} \qquad (55)$$

This is the force at P due to one turn ; the summation for the whole solenoid is now required.

The field strength at P due to a small segment of solenoid dx long is :—

$$h = \frac{2\pi r\,.\,i\,.\,\sin\theta}{r^2 + x^2}\,.\,N dx \text{ gauss}$$

where N is number of turns per unit length, *i.e.*, per centimetre.

But $$dx = AC = \frac{AE}{\sin\theta} \qquad \text{as } d\theta \to 0$$

$$= \frac{\sqrt{r^2 + x^2}}{\sin\theta}\, d\theta$$

$$\therefore \qquad h = \frac{2\pi r i}{\sqrt{r^2 + x^2}}\, N\, d\theta$$

$$= 2\pi N i \sin\theta \,.\, d\theta \text{ gauss} \tag{56}$$

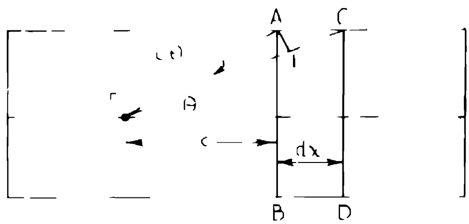


FIG 116 —Force on unit pole at centre of solenoid due to whole winding

If a long solenoid is considered then the limits of θ are 0 and π, so that

Resultant field strength at P is

$$H = \int_0^{\pi} 2\pi N i \times \sin\theta d\theta$$

$$= 2\pi N i \left[-\cos\theta\right]_0^{\pi}$$

$$= 4\pi N i \text{ gauss}$$

This is with i in EMU With I in amperes, the field strength at the centre of the solenoid is

$$H = \frac{4\pi N I}{10} \text{ gauss} \tag{57}$$

Note To obtain this result it was necessary to assume a *long* solenoid *i e* long compared with its radius If, in practice, this could not be assumed, then the limits of θ would need to be applied afresh to equation 56

Force on a conductor

So far the force on a magnetic pole due to a current has been considered, but if the conductor exerts a force on the magnet, then the magnet must exert an equal but opposite force on the conductor The direction of this force is shown in Fig 117, and is given by *Fleming's left-hand rule* —

Extend the thumb and first two fingers of the left hand mutually

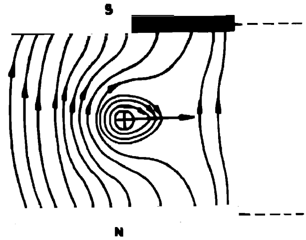


FIG. 117.—Force on current-carrying conductor in magnetic field.

at right angles. Place first finger in direction of field, second finger in direction of current; the thumb then gives the direction of the force acting on the conductor (see Fig. 118). The strength of this force can be calculated by applying Laplace's Law.

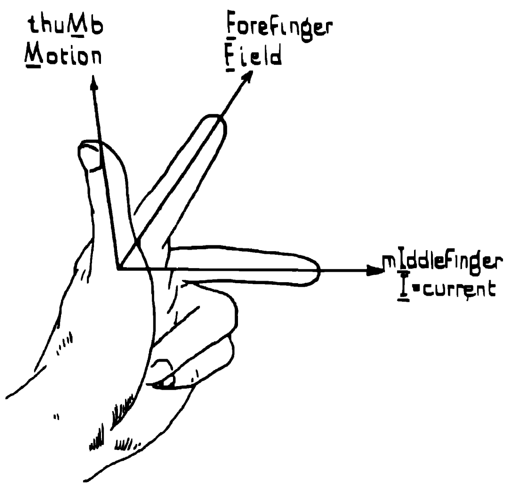


FIG. 118.—Fleming's left-hand rule.

Consider a magnetic pole of strength m, r cm. from a small segment of wire dl long, and carrying a current i EMU (*see* Fig. 119). The force on a pole m due to a current i is given by :—

$$F = \frac{m \cdot i \cdot dl}{r^2} \text{ dynes.}$$

The reaction on the conductor due to the magnetic pole must also be equal to $\frac{m \cdot i \cdot dl}{r^2}$ dynes.

But $\frac{m}{r^2}$ is the field strength at the conductor.

Therefore the force on a conductor of length dl in a field of strength H is $H \cdot i \cdot dl$ dynes.

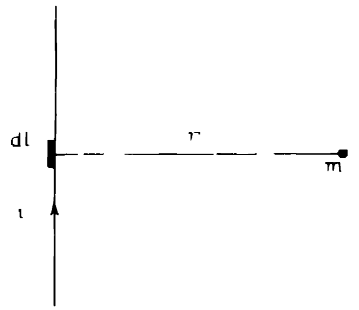


FIG. 119.—Pole distant r from current-carrying conductor.

If the field is uniform over the whole length l of the conductor, then :—

$$\text{Force} = \frac{H\,I\,l}{10}\ \text{dynes} \tag{58}$$

where I is in amperes.

Force on a coil in a magnetic field

Now consider a rectangular coil suspended in a uniform field, and carrying a current of I amps.

With the field into the paper the forces acting on the sides of the rectangle are as shown.

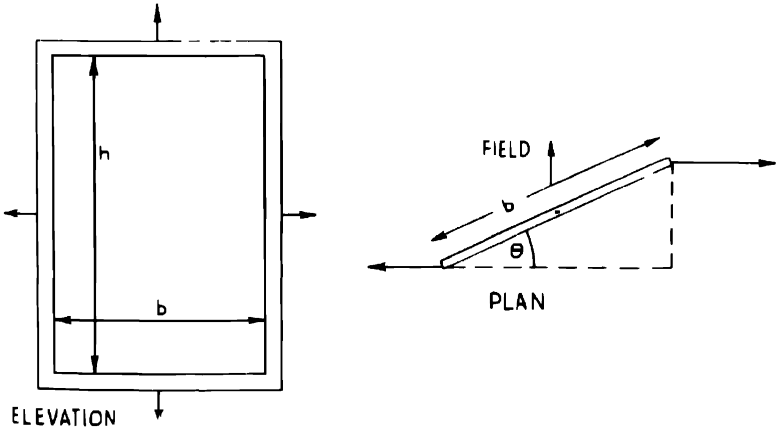


FIG. 120.—Rectangular coil in uniform magnetic field.

If the coil is now rotated through an angle θ about its vertical axis the two horizontal forces will form a couple which will tend to turn it back to its original position.

$$\text{Force on one conductor} = \frac{HIh}{10} \text{ dynes.}$$

$$\therefore \text{ Couple on coil} = \frac{HIh\,b \sin\theta}{10} \text{ dyne-cm.}$$

$$\text{But} \qquad hb = \text{area, } A\text{, of coil}$$

$$\therefore \text{ Couple on coil of } T \text{ turns} = \frac{HIAT}{10} \sin\theta \text{ dyne-cm.} \qquad (59)$$

Electromagnetic induction

Two convenient laws state the theory of electromagnetic induction very concisely :—

Faraday's Law.—*When the magnetic flux through a circuit is changing, an induced EMF is set up, and its magnitude is proportional to the rate of change of flux.*

Lenz's Law.—*The EMF induced in any circuit is always in such a direction that its effect tends to oppose the motion or change producing it.*

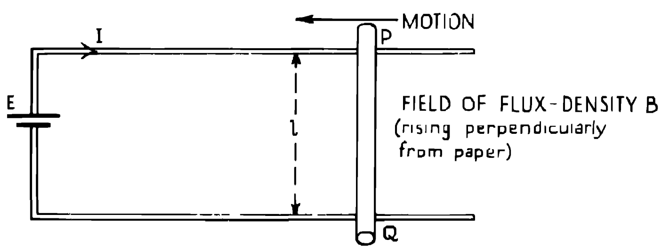


FIG. 121.—Bar moving in magnetic field

Consider the circuit shown in Fig. 121, composed of heavy-gauge conductors of negligible resistance, with a magnetic field having a flux-density B acting at right angles to the page.

Due to the force produced by the magnetic flux and the current i EMU through PQ, the bar PQ will be pulled along the parallel bars towards the left.

Let it move a distance dx in time dt.

Then the rate of change of flux-linkages is $B \,.\, l \,.\, \frac{dx}{dt}$

Now in moving the conductor, work is done.

$$\text{The force} = Bil \text{ dynes}$$

$$\therefore \text{ the work done} = Bil \,.\, dx \text{ ergs.}$$

This work is done by energy from the cell necessary to overcome the back EMF caused by motion.

Let the back EMF $= e$ EMU

Then the work done $= -ei\,dt$ ergs

$\therefore \quad Bil\,dx = -ei\,dt$

or
$$e = -Bl\,.\frac{dx}{dt} = -\frac{d\Phi}{dt}\text{ EMU} \qquad (60)$$

$\therefore$ the induced EMF in EMU $= -$ Rate of change of flux-linkages.

Hence the induced EMF in volts $= -$ Rate of change of flux-linkages $\times 10^{-8}$.

Fleming's right-hand rule

The direction of the induced EMF in such a case may be deduced by Lenz's Law. A convenient method of determining the direction of motion is given by Fleming's right-hand rule :—

> *Extend the thumb and first two fingers of the right hand mutually at right angles. Place first finger in direction of field, thumb in direction of motion ; the second finger will then give the direction of the induced EMF, and of the resulting current in a closed circuit (see Fig. 122).*

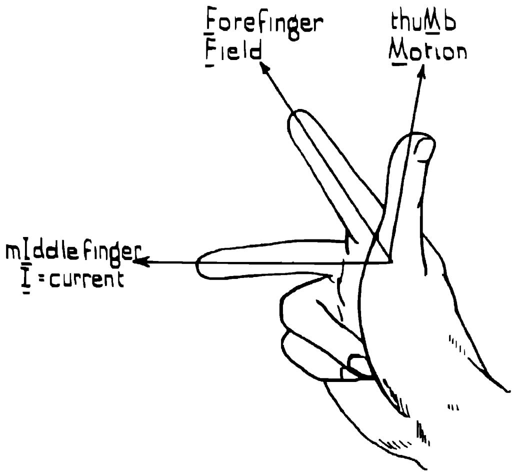


FIG. 122 – Fleming's right-hand rule.

Rotation of a coil in a uniform field

Consider a rectangular coil rotating at angular velocity ω radians per second, as shown in Fig. 123.

Then the velocity of one side of the coil is $\omega\frac{b}{2}$ as shown.

Velocity at right angles to field $= \omega\frac{b}{2}\sin\theta$ where $\theta = \omega t$, *i.e.* the angle turned through after time t.

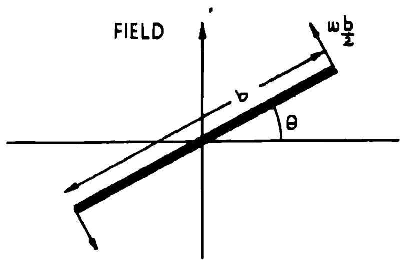


FIG. 123 —Coil rotating in uniform magnetic field.

$$\text{Rate of change of flux-linkages} = -2\left(\omega\frac{b}{2}\sin\theta\right)hB$$

$$= -\omega BA \sin\theta$$

where h is the height of the coil
B is the flux density
A is the coil area $= bh$.

Induced EMF in 1 turn $= -$ Rate of change of flux-linkages $\times$ 10^{-8} volts

$$= +\omega BA \sin\theta \,.\, 10^{-8} \text{ volts.}$$

Induced EMF in coil of T turns

$$= \omega BAT \sin\theta \,.\, 10^{-8} \text{ volts.} \qquad (61)$$

Thus it is seen that the EMF varies with $\sin\theta$. It will be zero, therefore, when $\theta = 0$, *i.e.* when the coil is at right angles to the field; and a maximum when θ is 90°, *i e.* when the coil is parallel to the field. The value of the EMF at the different positions is shown in Fig. 124 by a sine wave of amplitude $\omega BAT \,.\, 10^{-8}$ volts.

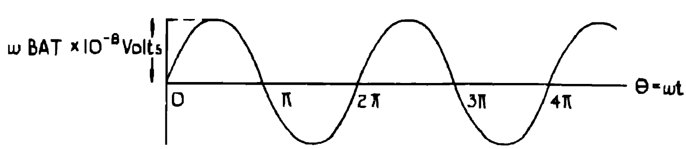


FIG 124.—Sinusoidal EMF induced into coil of Fig 123.

The instantaneous value of the EMF is seen from equation 61 to be $e = \omega BAT \sin\omega t \,.\, 10^{-8}$ volts. This is of the form :—

$$e = E_{max} \sin\omega t \qquad (62)$$

where $$E_{max} = \omega BAT \,.\, 10^{-8} \text{ volts.} \qquad (63)$$

The coil described above is the simplest form of " alternating current " generator, and equation 62 will be seen in Chapter 4 to be of fundamental importance in the study of AC.

Inductance

In the preceding sections induced EMFs due to change of flux in a coil have been considered. In each case the flux considered was due to some external field. When a current flows through a coil, however, it sets up a field of its own, and any change in this current alters this field. So without external aid a changing flux is obtained in the coil, and an induced EMF is therefore set up in such a direction as to tend to maintain the flux at its original density. This is termed "self-induction".

A coil has a self-inductance (L) of 1 henry when a change of current of 1 ampere per second induces a back EMF of 1 volt. Hence it follows that the back EMF is :—

$$e \text{ (in volts)} = -L \text{ (in henries)} \times \frac{dI}{dt} \text{ (in amps/sec.)} \qquad (64)$$

The negative sign is introduced because, when $\frac{dI}{dt}$ is positive, *i.e.*, when the current is increasing, the induced EMF e will be opposing the applied EMF.

From (64) : $$e = -L\frac{dI}{dt}$$

From (60) : $$e = -T\frac{d\Phi}{dt}10^{-8}$$

$$\therefore \quad L = T\frac{d\Phi}{dI}10^{-8}$$

$$\frac{d\Phi}{dI} = \frac{2\pi A\mu}{10r} \quad T = \frac{\Phi}{I} \qquad \text{(from eq. 54)}$$

$$\therefore \quad L = \frac{\Phi T}{I} \times 10^{-8} \text{ henries} \qquad (65)$$

i.e., self-inductance = flux-turns per ampere $\times 10^{-8}$.

Again, the field in the centre of a long solenoid is given by :—

$$H = \frac{4\pi TI}{10l} \text{ gauss}$$

$$\therefore \quad \text{Flux } \Phi = \frac{4\pi TI\,\mu A}{10l} \text{ lines}$$

and $$L = \frac{4\pi T^2\mu A}{10l} \times 10^{-8} \text{ henries} \qquad (66)$$

This last expression shows that the inductance of a long solenoid can be calculated from the purely physical dimensions :—

	Number of turns,	T
	Length,	l
	Permeability of core material,	μ
and	Cross-sectional area,	A

Mutual inductance

In the case of self-induction, the change of flux was caused by current variation in the same coil. This change of flux may,

however, be caused by current variations in a second coil that is linked with the first magnetically. The induced EMF is then due to "mutual induction".

Two coils have a mutual inductance (M) of 1 henry when a change of 1 ampere per second in one produces an EMF of 1 volt in the other. Hence it follows that the induced EMF is:—

$$e_2 \text{ (in volts)} = M \text{ (in henries)} \times \frac{dI_1}{dt} \text{ (amps/sec.)}.$$

As before, it can be shown that :—
Mutual inductance = Flux-turns in secondary per ampere in primary $\times 10^{-8}$

and $$M = \frac{4\pi T_1 T_2 \mu A}{10l} \times 10^{-8} \text{ henries}$$

For this to be true, however, the two coils must be fully linked magnetically. Otherwise the field, $\frac{4\pi TI}{10l}$, due to the primary, may not link with the whole of the secondary, and the derivation of the formula would be false.

Inductance of two parallel wires

The inductance of two parallel wires, radius a, distance between centres c, is given by :—

$$L = 0 \cdot 1609 + 1 \cdot 481 \log_{10} \frac{c - a}{a} \text{ mH per mile loop} \qquad (68)$$

Construction of inductances

For air-cored inductances, formulae can be used to determine the number of turns T required, but it is usually simpler in practice to use trial and error methods. Turns can be added or removed to alter the inductance L, remembering that L varies as T^2. The same methods apply to iron-cored inductances, although in this case the inductance can also be varied by adjusting the air-gap in the core.

Example.- Using a former $\frac{1}{2}$ inch square, length of winding $\frac{1}{2}$ inch, find how many turns are required for an inductance of 1 mH (air cored).

$$L = \frac{4\pi T^2 A}{10^9 . l} \text{ henries}$$

where l is the length in cm.

$$= 2 \cdot 54 \frac{4\pi T^2 A'}{10^6 . l'} \text{ mH}$$

where l' is the length in ins., and A' is the area in sq. ins.

In this case, $A' = \frac{1}{4}, l' = \frac{1}{2}, L = 1.$

Hence, $$1 = \frac{2 \cdot 54 \times 4\pi \times T^2 \times \frac{1}{4}}{10^6 . \frac{1}{2}}$$

$$\therefore \qquad T^2 = \frac{10^6 \, . \, 4}{2 \times 2{\cdot}54 \times 4\pi} = 251^2$$

251 turns are therefore required.

Such an inductance was constructed and found to have an inductance of 1·002 mH.

Circuits containing inductance and resistance

When the key is depressed in a circuit as shown in Fig. 125, the battery tries to drive a current of strength $\frac{E}{R}$ through the circuit, but owing to the inductance L the current is initially

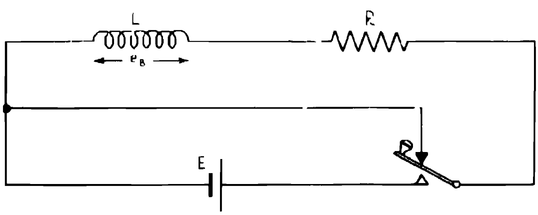


FIG. 125.– Application of LMF to inductance and resistance in series.

zero and slowly builds up to this value. The growth of current is comparable with the charging of a condenser discussed earlier.

Let the rate of change of current be $\frac{di}{dt}$

The back EMF $\qquad e_B = -L\frac{di}{dt}$

But $\qquad i = \frac{E + e_B}{R}$

$\therefore \qquad e_B = -E + iR$

Equating the two values of e_B :—

$$-E + iR = -L\frac{di}{dt}$$

$$\therefore \qquad \frac{di}{dt} + i\frac{R}{L} = \frac{E}{L}$$

Multiply by $e^{\frac{Rt}{L}}$:—

$$\frac{di}{dt} \cdot e^{\frac{Rt}{L}} + i \cdot \frac{R}{L} \cdot e^{\frac{Rt}{L}} = \frac{E}{L} \cdot e^{\frac{Rt}{L}}$$

$$\therefore \quad \frac{d}{dt}\left[i \cdot e^{\frac{Rt}{L}}\right] = \frac{E}{L} \cdot e^{\frac{Rt}{L}}$$

$$\therefore \quad i.e^{\frac{Rt}{L}} = \frac{E}{R} \cdot e^{\frac{Rt}{L}} + K$$

(where K is a constant)

$$\therefore \quad i = \frac{E}{R} + K \cdot e^{-\frac{Rt}{L}}$$

When $t = 0$, $i = 0$, hence K may be determined :—

$$K = -\frac{E}{R}$$

$$\therefore \quad i = \frac{E}{R}\left(1 - e^{-\frac{Rt}{L}}\right) \qquad (69)$$

This, then, is the equation for the graph shown in Fig. 126.

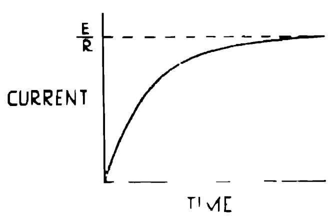


FIG. 126.—Graph showing growth of current in L–R circuit.

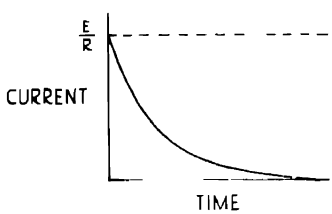


FIG. 127.—Graph showing decay of current in L–R circuit.

If the key in the above circuit is now released, the current tends to cease, but it is partially maintained by the back EMF, so that it dies away exponentially.

$$\text{Again} \quad e_B = -L\frac{di}{dt}$$

$$\text{But} \quad i = \frac{e_B}{R}$$

$$\therefore \quad e_B = iR$$

$$\therefore \quad iR = -L\frac{di}{dt}$$

$$\therefore \quad \frac{di}{i} = -\frac{R}{L}dt$$

Integrating :—

$$\log_e i = -\frac{Rt}{L} + K$$

or $$i = Ke^{-\frac{R}{L}t}$$

When $t = 0$, $$i = \frac{E}{R}$$

$\therefore$ $$K = \frac{E}{R}$$

Thus $$i = \frac{E}{R} \cdot e^{-\frac{R}{L}t} \tag{70}$$

and this is the equation for the graph of Fig. 127 showing the gradual decrease of current.

Time constant

The time-constant of a circuit containing inductance and resistance is equal to $\frac{L}{R}$. If L is expressed in henries and R in ohms, $\frac{L}{R}$ is equal to the time in seconds for the current to reach 63 per cent. of its final value when the circuit is closed; or for the current to fall to 37 per cent. of its initial value when the circuit is broken. (Compare this with the time-constant CR for a circuit containing inductance and capacity, CR being the time for the *charge* on the condenser to reach 63 per cent. of its final value, or fall to 37 per cent. of its maximum value.)

DC METERS

Meters may be roughly divided into two classes :—

(1) Voltmeters,

(2) Ammeters.

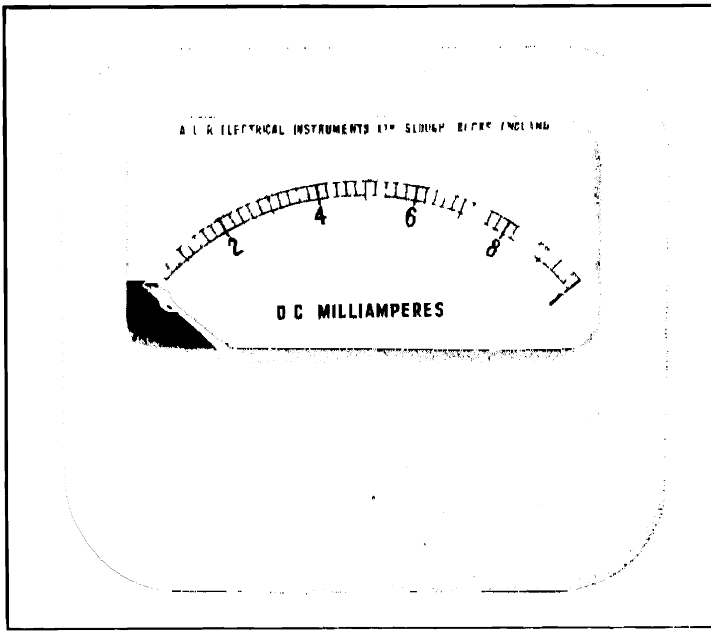
A voltmeter is always placed in parallel with the circuit, being connected to the two points between which the potential difference is to be measured. Therefore, to prevent disturbance of current distribution in the circuit, the voltmeter must have a high resistance. Ammeters, on the other hand, are placed in series with the circuit, and, for a similar reason, must have a low resistance.

Most meters are designed as sensitive milliammeters, and adapted as voltmeters and ammeters by the inclusion of series resistances or parallel shunts.

Moving coil meters

The principle underlying the operation of this type of meter is that expounded in the last section. A coil is suspended in a strong magnetic field, and, when a current I flows, it is turned by a couple of strength

$$\frac{HIAT \sin \theta}{10} \text{ dyne-cm.}$$



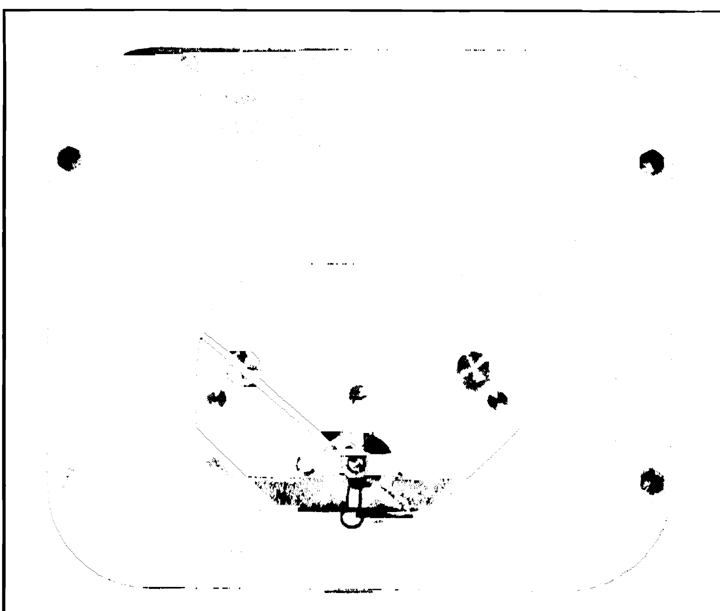


PLATE 5.—Moving coil meter

The field is supplied by a large permanent magnet as shown in Fig. 128, and, by placing a soft iron cylinder between the pole pieces, which are themselves shaped, this field is made radial. This means that whatever the position of the coil, $\sin \theta = 1$, and the turning couple is always constant and of value equal to $\frac{HIAT}{10}$ dyne-cm.

A restoring couple is supplied by a phosphor-bronze suspension or a spring. The latter is the more common in the normal portable

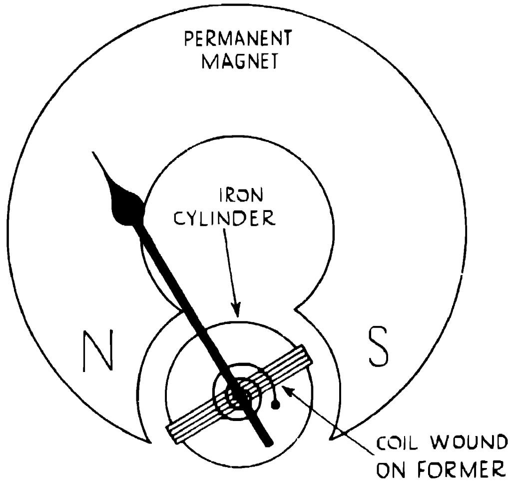


FIG 128.—Moving coil meter.

type of meter. This restoring couple is directly proportional to the deflection, *i.e.* $-k\theta$, where k is a constant. Therefore at equilibrium position

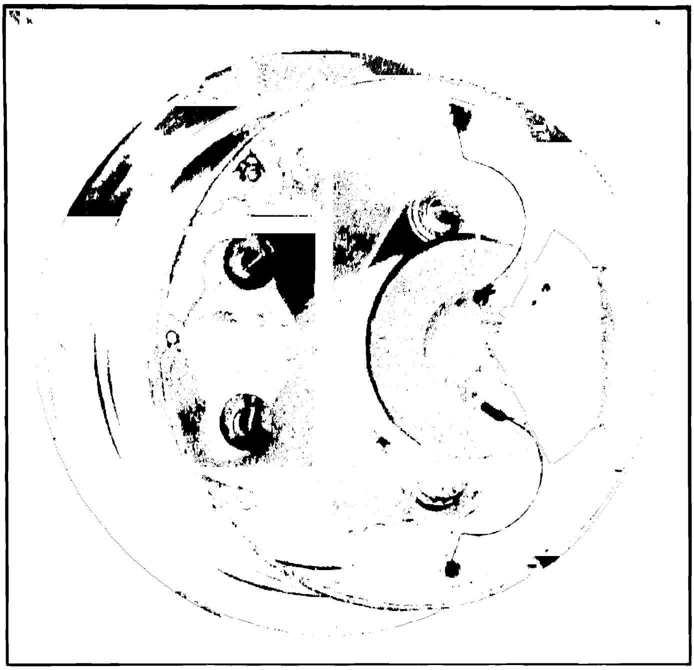
$$k\theta = \frac{HIAT}{10}$$

$$\therefore I = \frac{10 \, . \, k}{HAT}\theta$$

$$= K \, . \, \theta$$

The deflection is thus proportional to the current, so giving a linear scale.

The coil gains momentum as it swings towards this position and, to prevent this momentum causing unwanted oscillation about the equilibrium point, " damping " is introduced by winding the coil on a metal former. As the coil rotates, eddy currents are induced into this former, using up the kinetic energy of the system. When the coil reaches its equilibrium position it has no energy to swing further, and the turning and restoring couples are balanced, so that efficient damping is obtained.



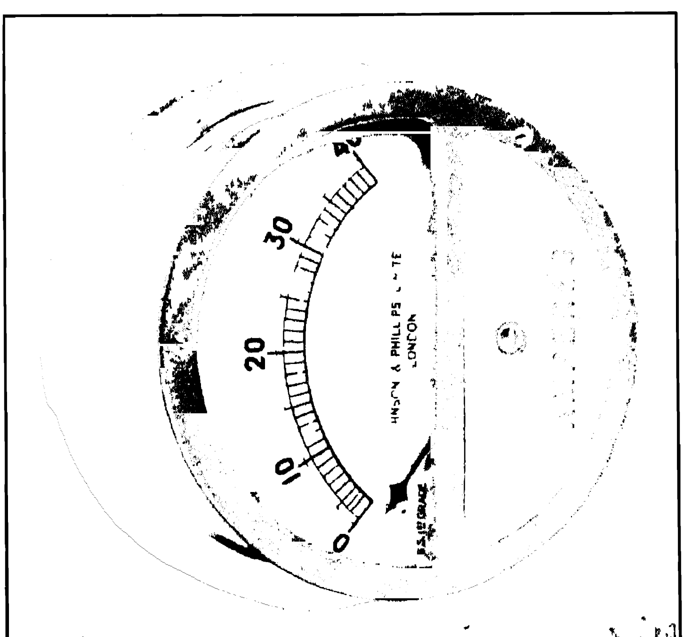


PLATE 6.—Moving iron meter—repulsion type.

"Ballistic" meters discard this damping and the coil is wound on a light non-metallic former They are used to measure discharge or similar short-duration currents, when the initial swing of the meter is proportional to the quantity of electricity passed

The moving coil milliammeter may be used as a voltmeter when a series resistance is connected For use as an ammeter a shunt must be connected in parallel with the meter

The passage of current through the coil raises its temperature and alters its resistance It is normal therefore to fit a "swamp" resistance in series with the coil The resistance being of a metal having a low temperature coefficient such as manganin varies little and thus reduces the effect of temperature on the meter as a whole

In the case of an ammeter the meter and swamp are connected in series, and the shunt is connected in parallel across both

Moving iron meters

This type of meter is usually made in one of two designs, the first is illustrated in Fig 129 The current to be measured is passed through the coil and sets up a magnetic field This magnetises the soft iron and draws it into the centre of the coil.

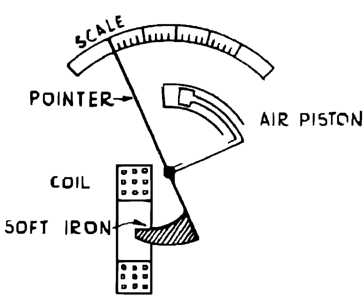


Fig 129 —Attraction type moving iron meter

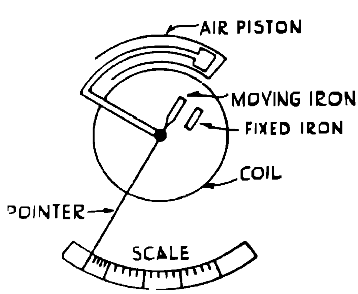


Fig 130 —Repulsion type moving iron meter

The force acting on the iron is dependent on the magnetism of the iron and on the field of the coil both of which are proportional to the current so that it is not directly proportional to the current but is proportional to the square of the current Therefore, unlike that of the moving coil meter the scale is not linear, but an approximately linear scale can be produced by careful shaping of the iron

The restoring couple is supplied by a spring or sometimes by gravity The damping is almost invariably obtained by the use of an air piston or "dashpot"

The second type, illustrated in Fig 130, is often known as the "repulsion" type meter.

PLATE 7.—Hot wire meter

Two iron bars are situated axially in a short solenoid. One is fixed, while the other is movable and attached to a pivot that also carries the pointer. When current flows through the coil the bars are equally magnetised and repel each other. The repulsion gives a deflection on the scale which is calibrated for direct reading. The repulsion is proportional to the product of the pole strengths of each magnet and therefore to I^2, giving again a non-linear scale.

Similar restoring and damping systems are used as in the previous type.

By varying the type of wire used in the coil the resistance of the meter can be varied, and so both voltmeters and ammeters can be made without need for any shunt or swamp resistances.

Both types of moving iron meters are susceptible to stray magnetic fields, and, since these would naturally cause deflection, good screening is necessary. Hysteresis also affects the readings, in that a higher reading will be given for decreasing currents than for increasing currents, while the retentivity will give a small deflection when no current is flowing. These defects are now reduced by the use of the alloy "mumetal", which has a hysteresis loss small enough to be neglected. By these means the moving iron meter can be made very accurate, and, owing to its robustness, it is more suitable for certain purposes than the moving coil meter. A further advantage is that the deflection is proportional to the square of the current, and this fact enables it to be used for AC as well as DC (*see* Chapter 6).

Hot-wire meters

Passage of current through a resistance wire causes generation of heat, and the wire expands with the resultant temperature rise.

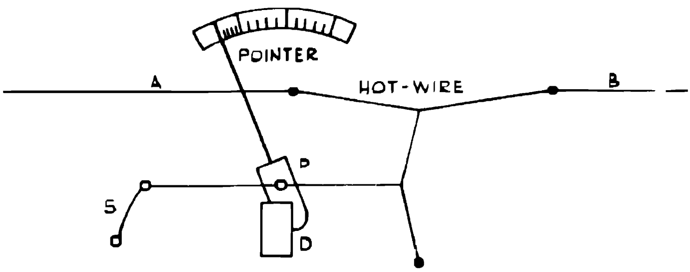


FIG. 131 Hot wire meter.

It can be shown that the expansion is proportional to the square of the current flowing, and this is the principle adopted in hot-wire meters, though due to practical factors the increase in length is not exactly proportional to I^2. Each meter must therefore be individually calibrated.

Fig. 131 shows the essential features. The current to be measured passes through the hot-wire from A to B. The hot-wire

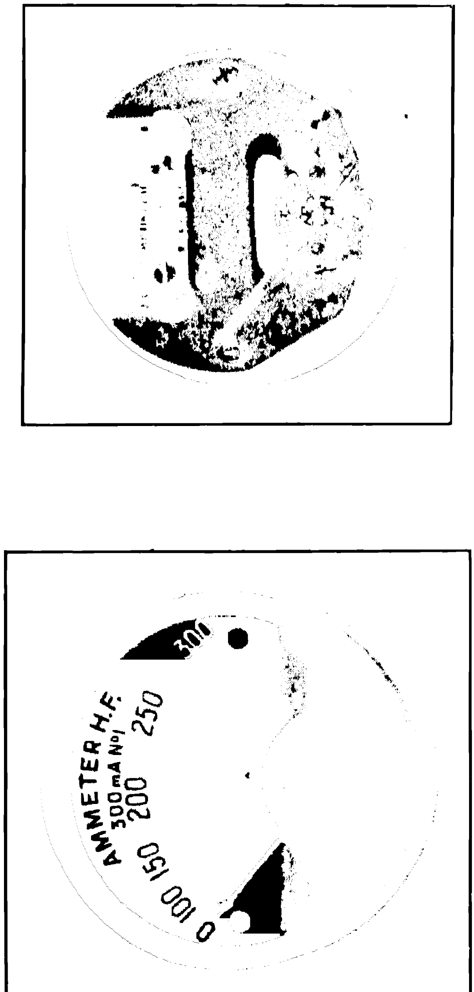


PLATE 8.—Thermo-couple meter.

is usually made of manganin, since the resistance of this metal varies little over the temperature range employed. The resultant sag in the hot-wire is taken up by tension applied from a phosphor-bronze spring S, through a silk thread that passes round pulley *P* and is attached to the top of *S*. Movement of the silk thread, due to change in the length of the hot-wire, rotates the pulley, and the scale is calibrated to read the current directly.

Damping is provided by a light aluminium plate, attached to the pointer, passing between the poles of a permanent magnet *D*. Contraction of the wire should return the meter to zero after use, but in practice this is not always the case and adjustments to the hot-wire have to be made. This fact, together with its high power consumption, its mechanical frailness and its liability to damage from overload renders its use unsuitable in many cases; but, like the moving iron meter, it has the advantage that it can be used for both DC and AC measurements.

Thermo-couple meters

The basic fact used in the construction of these meters was discovered as far back as 1806 and is known after the discoverer as the Seebeck effect.

If a circuit comprised of different metals is at one temperature throughout, there is no resultant EMF. If, however, one junction between two dissimilar metals is at a different temperature from

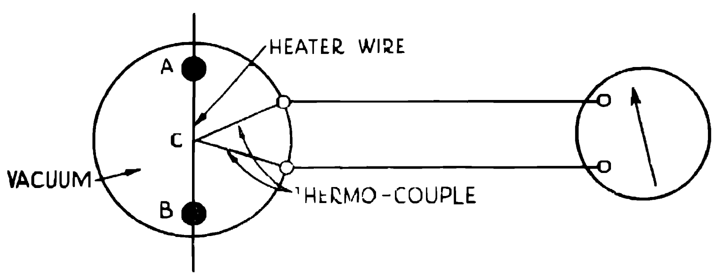


FIG. 132.—Thermo couple meter

the remainder of the circuit, an EMF is set up and a current will flow; this EMF is known as a "thermo-electric" EMF. This phenomenon can easily be demonstrated, using two dissimilar metals such as copper and iron, but the usual "thermo-electric couple", as it is called, used in such meters consists of bismuth and antimony, as this combination gives a large EMF, although numerous other combinations may be used.

As the temperature of the thermo-couple is increased, the EMF at first increases; but, after a certain rise in temperature, it ceases to increase and, finally, decreases until it reaches zero and starts to build up again with reversed polarity. It follows that only the

first portion (that of EMF increasing with temperature) can be used for thermo-electric meters

The construction of the meter is shown diagrammatically in Fig 132

The current to be measured passes between A and B, raising the temperature of the heater wire Attached to the centre of AB is the thermo-junction and, as the temperature of this junction increases so the EMF across the couple rises, the resultant current is passed through an ordinary DC meter usually of moving coil type The meter is calibrated to read directly the current flowing in the external circuit of which the heater wire AB forms a part

This type of meter is suitable for both AC and DC It is so built that the couple is neatly encased in the ammeter case, and it often resembles an ordinary moving coil meter though as it depends for operation on heating effect its scale is non-linear

Shunt and series resistances

If a meter has been designed as a sensitive milliammeter the use of shunts (in parallel) and series resistances is essential if the meter is to be used as an ammeter or voltmeter

The principle is as follows if a meter reads up to 100 mA and it is required to use it for measurements up to 1 ampere, then the inference is that, when 1 ampere passes through the circuit, only 100 mA must pass through the meter This will give full-scale deflection, and recalibration will enable direct reading of the currents to the higher limit Only 100 mA pass through the meter, therefore 900 mA must pass through shunt Thus, the required shunt $= \frac{1}{9}G$, where G is the resistance of the meter.

The principle of voltmeter and series resistance is similar, and is illustrated in the following example —

Q A meter of resistance 40 ohms gives a deflection of one scale division for a current of 1 milliampere Find the series resistance required to change it into a voltmeter reading 1 volt per scale division

A Voltage across meter for 1 scale division deflection $= \frac{40}{1{,}000}$ V.

But this must be the voltage across the meter when 1 volt is applied across the meter and series resistance,

$$\therefore \text{ Series resistance} = 40 \times \frac{\text{voltage across series resistance}}{\text{voltage across meter}} \text{ ohms}$$

$$= 40 \times \frac{1{,}000 - 40}{40} \text{ ohms} = 960 \text{ ohms.}$$

Some meters are provided with a set of shunts and series resistances so that their ranges and uses can be varied. As an

example, consider the meter shown in Fig. 133*a*, having an internal resistance of 200 Ω, and giving full-scale deflection when the current flowing through it is 0·75 mA; this corresponds to a voltage of 150 mV across the meter.

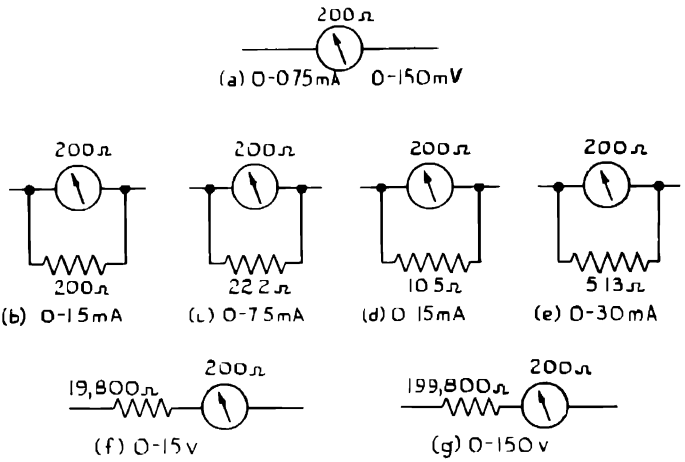


Fig. 133—Multi-purpose meter.

The range may be extended as follows:—

0—1·5 mA	using a shunt of 200 Ω
0—7·5 mA	using a shunt of 22·2 Ω
0—15 mA	using a shunt of 10·5 Ω
0—30 mA	using a shunt of 5·13 Ω
0—15 V	using a series resistance of 19,800 Ω
0—150 V	using a series resistance of 199,800 Ω.

CIRCUITS CONTAINING INDUCTANCE, CAPACITY AND RESISTANCE

It is interesting to consider the behaviour of the current in the circuit of Fig. 134 after the key is closed. It will be seen that there are three distinct cases, depending on the relative values of R, L and C. A differential equation is obtained by writing down an equation for the current t seconds after closing the key.

The voltage across C is $\frac{q}{C}$, and has polarity as shown.

At $t=0$, $q=0$, since the condenser is initially discharged, and $i=0$.

The voltage across L is $-L\frac{di}{dt}$, measured in the direction of i, and therefore Kirchhoff's Law gives:—

$$E - L\frac{di}{dt} - \frac{q}{C} = iR \tag{71}$$

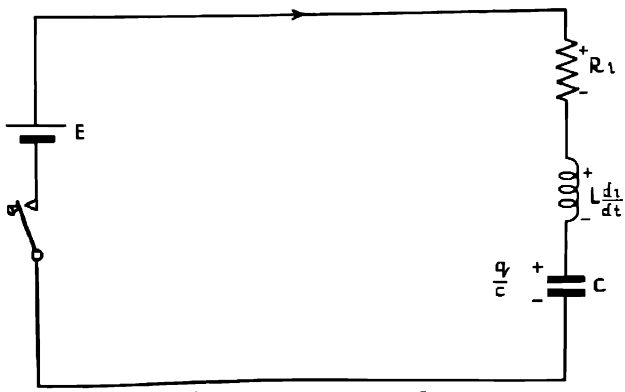


FIG 134 —Application of LMF to R, L and C in series.

To eliminate the variable q from the equation, differentiate both sides with respect to t and put $\frac{dq}{dt} = i$, giving :—

$$-L\frac{d^2i}{dt^2} - \frac{i}{C} = R\frac{di}{dt}$$

or

$$\frac{d^2i}{dt^2} + \frac{R di}{L dt} + \frac{i}{LC} = 0 \qquad (72)$$

The solution of this equation is :—

$$i = Ae^{m_1 t} + Be^{m_2 t},$$

where m_1 and m_2 are the roots of the equation :—

$$m^2 + m\frac{R}{L} + \frac{1}{LC} = 0$$

The roots are :—

$$m = -\frac{R}{2L} \pm \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}}$$

That is, $m = -a \pm b$, where $a = \frac{R}{2L}$ and $b = \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}}$

Hence, the general solution to equation 72 is :—

$$i = Ae^{(-a+b)t} + Be^{(-a-b)t} \qquad (73)$$

Case 1.—

$$\frac{R^2}{4L^2} > \frac{1}{LC} \qquad \therefore\ b \text{ is real.}$$

Equation 73 gives :—

$$i = e^{-at}(Ae^{bt} + Be^{-bt})$$
$$= e^{-at}(F\cosh bt + G\sinh bt) \qquad (74)$$

To determine the value of F and G, put $t = 0$.

This must give $i = 0$, hence $F = 0$.

Also, from equation 71 .—

$$\frac{di}{dt} = \frac{E}{L} \text{ when } t = 0$$

Differentiating equation 74 with $t = 0$ and $F = 0$

$$\frac{E}{L} = \left(\frac{di}{dt}\right)_{t=0} = bG$$

$$\therefore \qquad G = \frac{E}{bL}$$

Equation 74 now becomes :—

$$i = \frac{E}{bL} \epsilon^{-at} \sinh bt \qquad (75)$$

The way in which i varies with t is shown in Fig 135.

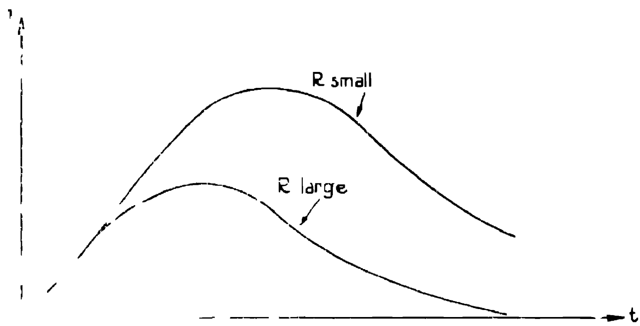


Fig 135 Current in circuit of Fig 134 when $\dfrac{R^2}{4L^2} > \dfrac{1}{LC}$.

Case 2 —

$$\frac{R^2}{4L^2} = \frac{1}{LC} \qquad \therefore\ b = 0$$

The solution of equation 72 in this case is .—

$$i = (A + Bt)\,\epsilon^{-at}$$

When $\quad t = 0,\ i = 0 \quad \therefore\ A = 0$

Hence $\quad i = Bt\epsilon^{-at}$

As before, when

$$t = 0,$$

$$\frac{di}{dt} = \frac{E}{L}$$

$$\therefore \qquad B = \frac{E}{L}$$

Thus, $$i = \frac{Et}{L} e^{-at} \qquad (76)$$

This curve is of similar shape to those in Fig. 135

Case 3.—

$$\frac{R^2}{4L^2} < \frac{1}{LC} \qquad \therefore b \text{ is imaginary}$$

Let $b = j\omega$, so that $\omega = \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}$

Equation 73 gives :—

$$i = \epsilon^{-at} (F \cos \omega t + G \sin \omega t)$$

and, as before, by considering $t = 0$, F and G may be evaluated, giving :—

$$i = \frac{E}{\omega L} \epsilon^{-at} \sin \omega t$$

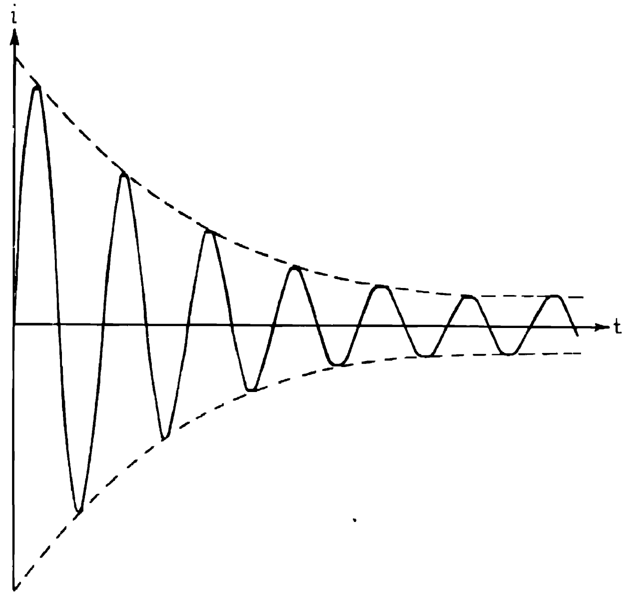


FIG. 136.—Current in circuit of Fig. 134 when $\frac{R^2}{4L^2} < \frac{1}{LC}$.

This shows that i has an exponentially decaying sinusoidal waveform, as shown in Fig. 136. Damped oscillations occur in this circuit at a frequency f given by :—

$$f = \frac{1}{2\pi} \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}$$

This is known as the "natural resonant frequency" of the circuit. The rate of damping is proportional to a, *i.e.*, $\frac{R}{2L}$, and will be least when R is small.

CHAPTER 4

ALTERNATING CURRENTS

DC theory deals with the behaviour of currents and voltages that are constant in magnitude and direction; AC theory deals with currents and voltages that are not constant, but that vary through some particular cycle of values which is repeated continuously at some fixed rate. The rate of repetition is called the "frequency" and is measured in cycles per second. Fig. 137 shows the graph of such a waveform; the vertical axis represents voltage, and the horizontal axis, time

It can be seen that the complete voltage cycle extends from *A* to *B*, where it is repeated. The duration of each cycle in this case is $\frac{1}{100}$th sec., and hence there are 100 cycles in one second, *i.e.*, the "frequency" is 100 cycles/second. The frequency may

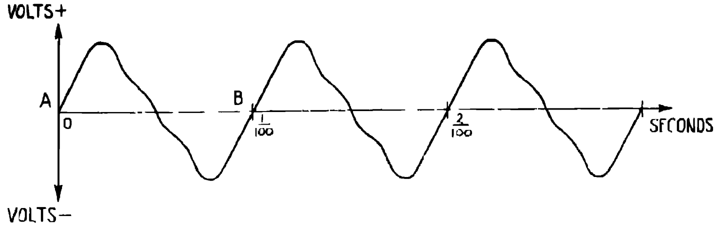


FIG. 137.—Typical alternating current waveform.

have any value up to infinity. For convenience this range can be roughly divided up into bands. Frequencies below 100 cycles/second are normally used for commercial power distribution; for example, in this country the supply frequency is usually 50 cycles/second. The next band is the "audio" frequency band, *i.e.*, the range of the human ear. The exact limits of this vary with individuals, but lie between 20 c/s and 20 kc/s (20,000 cycles/second); many people cannot hear above 10 kc/s. The frequency determines the "pitch" of a note; for example, middle "C" on a piano is taken, using the scientific scale, as 256 cycles/second. An increase in pitch of one octave is equivalent to *doubling* the frequency. Faithful reproduction of speech or music would demand a system that could work from 20 cycles/second to 20 kc/s; it is possible, however, to obtain satisfactory intelligibility of speech if frequencies between 300 and 2,000 cycles/second are reproduced. Above this range are

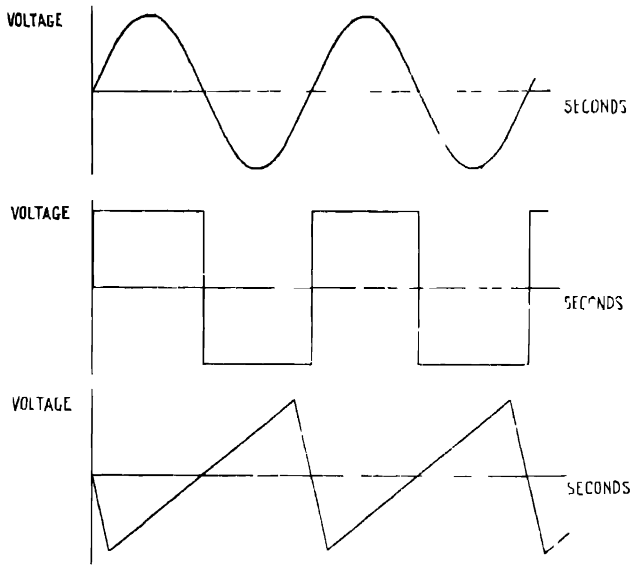


Fig. 138—Alternating voltages having (a) sinusoidal (b) square, and (c) triangular waveforms

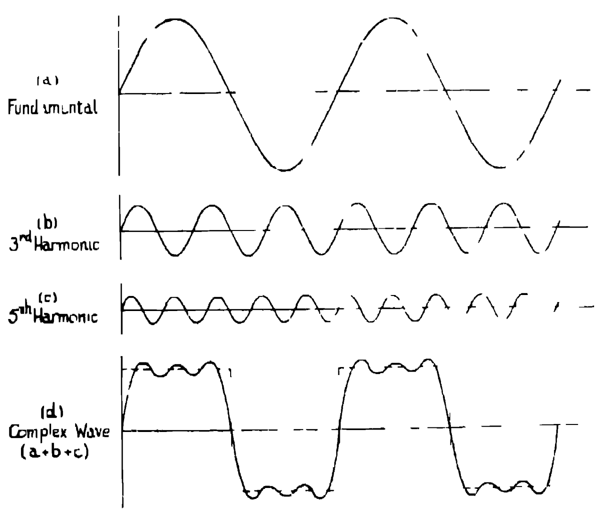


Fig. 139—Addition of fundamental sine wave, with third and fifth harmonics, to give an approximation to a square waveform.

the frequencies used for carrier telephony and wireless—up to several thousand megacycles—and these are further sub-divided

The shape or 'waveform' of an alternating voltage or current is just as important as the frequency Fig 138 shows three voltages that have the same frequency but have different waveforms

Since any recurrent waveform may have to be dealt with, the problem is to develop a general theory that can be applied to any particular case and Fourier's theorem provides the solution

Fourier's theorem

The mathematical aspect of Fourier's theorem has been dealt with in Chapter 2 In words this theorem states that *any recurrent waveform of frequency f can be resolved into the sum of a number of sinusoidal waveforms having frequencies f 2f 3f, The number of sine waves may be finite or infinite*

An alternative way of stating the theorem is to state that any steady note can be split up into a fundamental and harmonics" The fundamental has frequency f the second harmonic has frequency $2f$ the third harmonic has frequency $3f$ and so on Sounds produced by the human voice or musical instruments nearly always contain a large number of harmonics

This theorem thus reduces any waveform to a number of sine waves The square wave for example consists of a fundamental and all the *odd* harmonics up to infinity suppose the amplitude of the fundamental is 1 then the amplitude of the third harmonic will be $\frac{1}{3}$, that of the fifth will be $\frac{1}{5}$ and so on Fig 139 shows the result of taking up to the 5th harmonic it will be seen that this gives a good approximation to a square wave

The general theory developed in this chapter will therefore be based on the assumption that the waveform is sinusoidal The behaviour of other types of waveform can be investigated by applying Fourier's theorem and the Superposition theorem (see Chapter 5)

SINUSOIDAL WAVEFORMS

Fig 140 shows the graph of $e = E_{max} \sin 2\pi ft$ This is known as a sinusoidal waveform or sine wave and it has been shown in

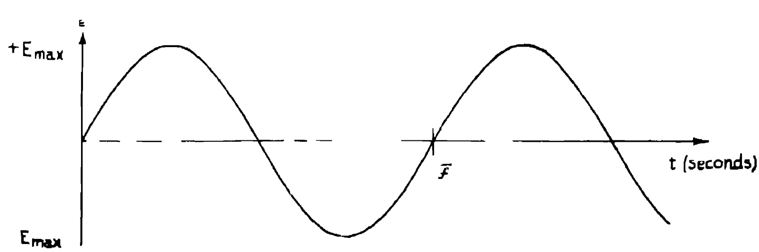


FIG 140—Graph of $e = E_{max} \sin 2\pi ft$

Chapter 3 that it is the form of voltage produced in a loop of wire rotated with a constant speed of rotation in a uniform magnetic field (the simple alternator).

From the equation it can be seen that the first cycle ends when

$$2\pi ft = 2\pi \text{ (radians)}$$

that is, after a time $t = \frac{1}{f}$ second.

Hence the "periodic time" of one cycle is $\frac{1}{f}$ seconds, and therefore there are f cycles per second, i.e. f is the frequency.

It can also be seen (remembering that the maximum value of $\sin\theta$ is 1) that the maximum value of e is E_{max}. Hence

$$e = E_{max} \sin 2\pi ft$$

is the equation of a sinusoidal EMF whose peak value is E_{max} and whose frequency is f. e represents the value of the EMF at any time t and is known as the instantaneous value. Similarly $i = I_{max} \sin 2\pi ft$ represents a sinusoidal current. The function need not necessarily be "$\sin 2\pi ft$", since "$\cos 2\pi ft$" or "$\sin(2\pi ft + \varphi)$", or "$A \sin 2\pi ft$" all represent sinusoidal waveforms.

Angular velocity

The voltage produced by a coil of wire rotating in a linear magnetic field has been shown to be $E_{max} \sin 2\pi ft$, where f is the speed of rotation in revolutions/second. If the angular velocity ω (in radians/sec) is taken as determining the speed, the equation becomes $e = E_{max} \sin \omega t$, for there are 2π radians to one revolution and hence f revolutions/second corresponds to $2\pi f$ radians/sec.

That is $$\omega = 2\pi f \qquad (1)$$

In AC problems it is often more convenient to deal with ω than with f, principally because the three symbols $2\pi f$ can be replaced by the single symbol ω.

At this stage it is worth mentioning certain approximations that can be made in calculations where extreme accuracy is not required. It will be found that $f = 800$ c/s gives $\omega = 5\,025$ radians/second, and that ω can normally be taken as $5\,000$, the error being only ½ per cent. Similarly $f = 1\,600$ c/s gives $\omega \simeq 10\,000$ radians/second, and so on. These approximations will be made in this chapter and are worth remembering.

Mean and RMS values

The mean value of any expression of the form $E_{max} \sin \omega t$ is zero over an indefinite period of time. It will be assumed when dealing with any waveform that the mean value is zero: otherwise it can be regarded as containing a DC component which must be dealt with separately.

It was shown in connection with DC (Chapter 3) that the power or heating effect of a voltage or current depended upon its *square*. This applies also to AC, but the square of a sine wave will vary from

instant to instant; it is, in fact, not the *instantaneous* value of the square that is important, but its *mean* value. It can be seen at once that this mean value will not be zero, for a *square* is always positive. This mean value is most conveniently expressed in the form of a voltage E such that E^2 is equal to the mean value of e^2. In this case E is the square root of the mean value of e^2, and is therefore called the *Root Mean Square* (RMS) value of the alternating voltage; it is that steady voltage which would give the same mean power effect as the alternating voltage. A similar definition applies to RMS values for alternating currents. These values are written as E_{RMS} and I_{RMS}. This is to distinguish these values from maximum values (E_{max} and I_{max}) and instantaneous values (e and i).

The value of E_{RMS} will now be calculated in terms of E_{max}. The first step is to calculate the mean value of e^2; this can be done by calculus, or more simply as follows:—

$$e = E_{max} \sin \omega t$$
$$\therefore \; e^2 = E_{max}^2 \sin^2 \omega t$$

The mean value of $\sin^2 \omega t$ is not obvious, but a useful trigonometric formula gives:—

$$\sin^2 \theta = \tfrac{1}{2}(1 - \cos 2\theta)$$

Applying this formula:—

$$e^2 = E_{max}^2 \cdot \tfrac{1}{2}(1 - \cos 2\omega t)$$
$$= \frac{E_{max}^2}{2} - \frac{E_{max}^2}{2} \cos 2\omega t$$

Considering the right-hand side, the first term is constant and its mean value is equal to $\frac{E_{max}^2}{2}$. The second term has a sine waveform, and its mean value over a period of time is zero.

$$\therefore \text{ The mean value of } (e^2) = \frac{E_{max}^2}{2}$$

Now $$E_{RMS} = \sqrt{\text{mean value of } e^2}$$

$$\therefore \quad E_{RMS} = \frac{E_{max}}{\sqrt{2}} \qquad (2)$$

which gives the RMS value in terms of the peak value.

Similarly $$I_{RMS} = \frac{I_{max}}{\sqrt{2}} \qquad (3)$$

Alternating voltages and currents are usually measured by their RMS values; it is important to remember this, as all insulation must be able to withstand the *peak* voltage. For example, the normal 230 volt mains supply has a peak voltage of $\sqrt{2} \times 230 = 325 \cdot 3$ volts.

AC + DC—If both AC and DC flow through a circuit at the same time the RMS value of the resultant is $\sqrt{I_{DC}^2 + I_{RMS}^2}$ where I_{DC} is the DC component and I_{RMS} is the RMS value of the AC component. The proof is similar to that given above.

Mean and average values—The mean value of an expression of the form $e = E_{max} \sin \omega t$ (which has been shown to be zero over an indefinite period of time) is usually called the average value of e. The term mean value of e is frequently used to represent the mean value of $|e|$, which is the average [illegible] and can be shown to be equal to $\frac{2}{\pi} I_{max} = 0.637 I_{max}$.

Form factor—The ratio of the RMS value to the mean value of a current or voltage is called the form factor of the waveform in question. In the case of a sine wave

$$\text{Form factor} = \frac{0.707 I_{max}}{0.637 I_{max}} = 1.11$$

Phase

Two waveforms of the same frequency [illegible] reach similar points of [illegible] the same time. The amount by which they are out of step is called the difference in phase, and is just as important as the magnitude when comparing two waveforms. Phase difference is [illegible] which one wave is ahead of or behind the other. [illegible] for example

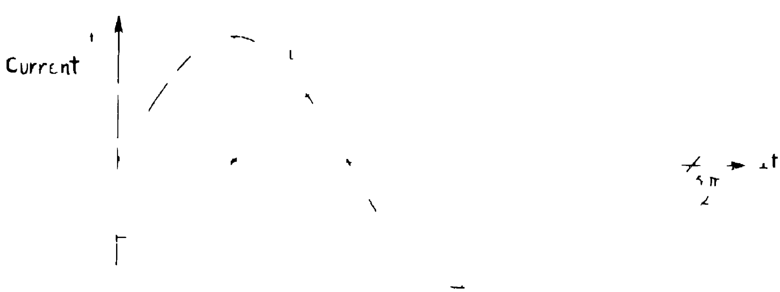


Fig. 141 Two waves 90° out of phase

the larger curve is 90° (or $\pi/2$ radians) ahead of the smaller one. This can be seen from the fact that the larger one is initially zero but the smaller one is zero and increasing in the same direction 90° later.

The difference in phase between two voltages or currents can easily be seen from the equations.

For example, $e = E_{max} \sin (\omega t + \varphi)$ represents a wave that is φ radians ahead of the wave $e = E_{max} \sin \omega t$

Note that the voltage e [illegible] $E_{max} \sin \omega t$ is 90° behind the voltage [illegible] $E_{max} \cos \omega t$ since $\cos \omega t$ [illegible] $\sin (\omega t + 90°)$.

Sum of two sine waves of same frequency

[illegible] of the same frequency is a [illegible]

[illegible]

(a) Draw [illegible] and add up point by point. The resultant will be [illegible] of the same frequency.

(b) [illegible]

[illegible]

Now any expression [illegible] can be written as

[illegible]

[illegible] $\sin (\omega t + \theta)$

[illegible] $\sin (\omega t + \theta)$ (4)

where θ [illegible]

This is a sine wave [illegible]

Representation by rotating vector

Neither the [illegible] method of representing [illegible] particularly when relative phase [illegible] It will [illegible] be shown that any sinusoidal [illegible] can be represented as part of a rotating [illegible]

Consider the [illegible] To calculate e from first principles the value of $\sin \omega t$ at any instant has to be found. This is done [illegible] by drawing the angle ωt anticlockwise from a starting line and dropping a perpendicular (*see* Fig. 142).

$$\text{The angle } \angle PON = \omega t \quad \text{Then } \frac{NP}{OP} = \sin \omega t$$

To calculate $e = I_{max} \sin \omega t$ make $OP = I_{max}$

Then $PN = I_{max} \sin \omega t$

$PN = e$

A moment later ωt will have altered but I_{max} is constant and therefore OP remains the same length. This is represented by Fig. 143.

In all cases e is equal to PN; if P falls below the horizontal axis e is negative.

If the figure is drawn for successive instants the line OP will rotate about O at a constant angular velocity ω. As both the length and the direction of OP are important it must be regarded

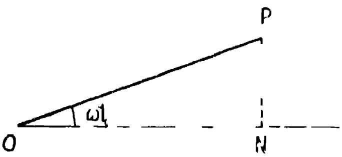


Fig. 142.—Representation of sinusoidal voltage by rotating vector OP

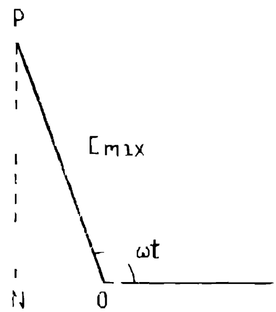


Fig. 143.—Representation of sinusoidal voltage by rotating vector OP

as a vector and e is equal to its vertical component. Note that e is only *part* of the vector. In practice it is easier to deal with a complete vector than with any particular part and it is for this reason that vectors are so important in A.C. theory. The reason is easy to see: e itself is a *varying* quantity but the complete vector has *constant* amplitude and *constant* angular velocity.

The representation of A.C. by vectors may be summarised as follows: any sinusoidal alternating E.M.F. of peak value I_{max} and angular velocity ω may be represented by the vertical component of a rotating vector whose length is equal to I_{max} and whose angular velocity is equal to ω. It can be seen that the frequency of the alternating voltage is equal to the speed of rotation of the vector measured in *revolutions per second*.

A similar definition applies to currents.

Note that the *horizontal* component could equally well have been taken.

Illustration of phase difference

Consider the voltage $e_2 = E_{2max} \sin (\omega t + \varphi)$. This is the vertical part of a vector whose length is E_{2max} and whose angle

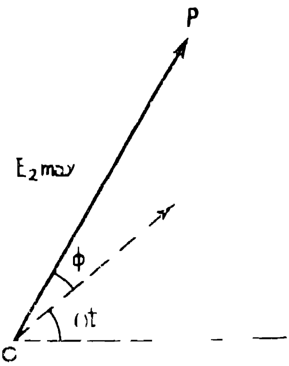


FIG 144 —Representation of sinusoidal voltage by rotating vector OP

is $\omega t + \varphi$ To draw the vector the angle ωt would be drawn as before and the constant angle φ added to it as in Fig 144

If this vector and the vector representing $e_1 = E_{1max} \sin \omega t$ are drawn on the same diagram for different values of ωt Fig 145 is obtained

Note that E_{1max} and E_{2max} remain in the same *relative* position, rotating together at the same speed In AC problems it is sufficient to consider vectors in one position only and to simplify calculation a position is usually chosen in which one of the vectors is horizontal Thus the two voltages would appear as in Fig 146

It is important to remember when dealing with these vector diagrams as they are called that although drawn in one fixed position the vectors are rotating at a constant angular velocity ω, and that the instantaneous voltages are the vertical components of the vectors at any instant The direction of rotation is anti-

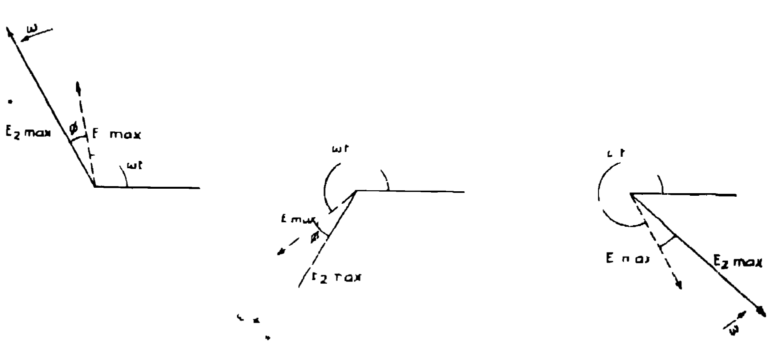


FIG. 145 —Representation of two sinusoidal voltages e_1 and e_2 by rotating vectors.

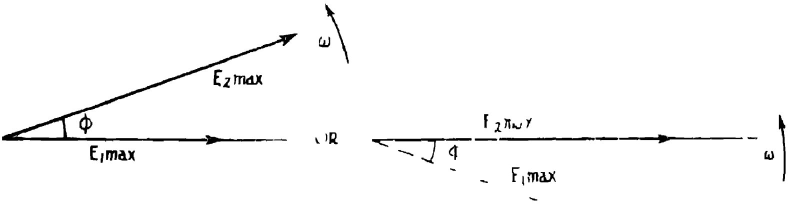


Fig. 146. Vector representation of the two voltages e_1 and e_2

clockwise, is shown in Fig. 146. It will be noted that a vector diagram is the best way of showing the phase difference between two waveforms.

Addition of voltages

The problem of addition and subtraction, already solved by two methods, is much simplified by using vectors. To find the sum of $e_1 = E_{1max} \sin \omega t$ and $e_2 = E_{2max} \sin (\omega t + \varphi)$ it is necessary to find the *sum of the vertical components of the two vectors*. This, of course, is equal to the *vertical component of the sum of the two vectors*; in other words, if the two vectors are drawn and added the resultant vector will represent the sum of the two voltages.

This is shown in Fig. 147.

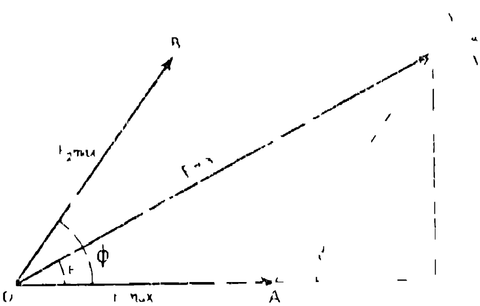


Fig. 147. Addition of two out of phase voltages by rotating vectors

Since E_{1max} and E_{2max} rotate together E_{max} will rotate at the same speed and the figure will not change its shape, hence E_{max} has the same frequency as E_{1max} and E_{2max}.

Applying the cosine formula to the triangle OAC:—

$$OC = \sqrt{OA^2 + AC^2 - 2 . OA . AC \cos (180^\circ - \varphi)}$$

i.e. $$E_{max} = \sqrt{E_{1max}^2 + E_{2max}^2 + 2E_{1max} E_{2max} \cos \varphi} \qquad (5)$$

Dropping a perpendicular CD on to OA,

$$\tan\theta = \frac{CD}{OA + AD}$$

$$\therefore \quad \theta = \tan^{-1}\frac{I_{2max}\sin\varphi}{I_{1max} + I_{2max}\cos\varphi} \tag{6}$$

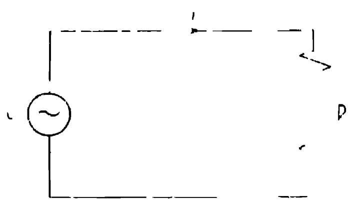
It will be seen from this that the use of vectors provides an easy way of determining the sum of two alternating voltages. Subtraction is done by subtracting the vectors. It is most important to note that (unless they happen to be in phase) alternating voltages *cannot be added directly*, that is, the peak or RMS values cannot be added. Addition or subtraction *must* be done vectorially.

AC CIRCUITS

The next problem is to consider the behaviour of alternating currents and voltages in various circuits consisting of combinations of inductance, capacity and resistance. It will be shown that in these circuits a sinusoidal voltage produces a sinusoidal current of the same frequency and *vice versa*. The current and voltage will not in general be in phase; the phase and magnitude relationships for particular circuits will now be determined.

Resistance, capacity and inductance

Resistance. Let a voltage $e = E_{max}\sin\omega t$ be applied to a resistance R (see Fig. 148).



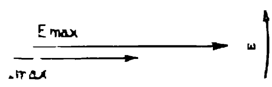


Fig. 148. Alternating voltage applied to a resistance. Fig. 149. Current through resistance in phase with the applied voltage.

Ohm's law applies at any instant; hence i can be determined from the equation $e = iR$ or $i = \frac{E_{max}}{R}\sin\omega t$.

Hence i is a sinusoidal alternating current in phase with e and having the same frequency. The peak value of i is $I_{max} = \frac{E_{max}}{R}$.

The two relationships are therefore:—

(*a*) e and i are in phase

$$(b)\ \frac{E_{max}}{I_{max}} = R \tag{7}$$

The vector diagram is therefore as shown in Fig. 149, where the two vectors have been separated for clarity.

Inductance.—Let $e = E_{max} \sin \omega t$ be applied to an inductance L (*see* Fig 150) The resistance of the inductance is taken as zero.

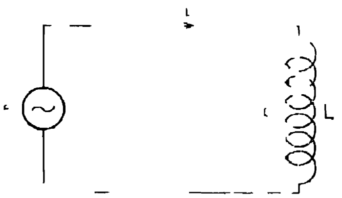


Fig 150 Alternating voltage applied to an inductance

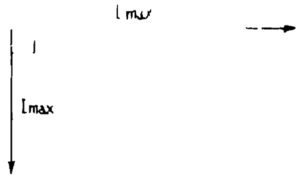


Fig 151 Current through inductance lagging 90° behind applied voltage.

The EMF induced across the inductance e' is given by $-L\frac{di}{dt}$.

Applying Kirchhoff's second law

$$e + e' = 0$$

$$\therefore \quad e - L\frac{di}{dt} = 0$$

$$\therefore \quad \frac{di}{dt} = \frac{e}{L} = \frac{E_{max}}{L}\sin \omega t$$

Integrating with respect to t

$$i = -\frac{E_{max}}{\omega L}\cos \omega t$$

This can be written as

$$i = \frac{E_{max}}{\omega L}\sin\left(\omega t - \frac{\pi}{2}\right)$$

which shows that

(*a*) i is 90° behind e

(*b*) $I_{max} = \dfrac{E_{max}}{\omega L}$. (8)

The vector diagram is shown in Fig 151

Capacity.—Let $e = E_{max} \sin \omega t$ be applied to a condenser of capacity C (*see* Fig 152)

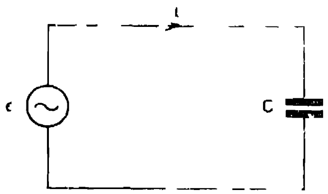


Fig 152 Alternating voltage applied to condenser

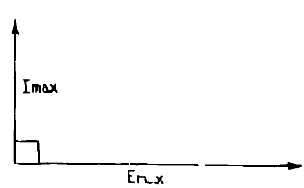


Fig 153 —Current through condenser leading the applied voltage by 90°

Measured *in the direction of* i, $e' = -\frac{q}{C} = -\frac{1}{C}\int i\,dt$

(Note the minus sign: if i is positive, e' is in opposition to i.)

Applying Kirchhoff's second law:—

$$e + e' = 0$$

$$\therefore \quad e = \frac{1}{C}\int i\,dt$$

$$\therefore \quad \int i\,dt = CE_{max} \sin \omega t$$

Differentiating with respect to t,

$$i = \omega CE_{max} \cos \omega t$$

This may be written as:—

$$i = \omega CE_{max} \sin\left(\omega t + \frac{\pi}{2}\right)$$

showing that:—

(a) i is 90° ahead of e.

(b) $I_{max} = \omega CE_{max}$. (9)

The vector diagram is shown in Fig. 153

Current-voltage relationships in AC circuits

All DC theory, including, in particular, Kirchhoff's laws, is based upon the fact that $E \div I$ is a constant and is equal to the resistance R. A similar relationship is required for AC if the same methods are to apply. But, taking instantaneous values, $e \div i$ will vary from 0 to $\pm\infty$ if the two are not in phase, so the law does not hold in its original form.

Consider, however, the result of dividing the *vectors* representing e and i. Denoting these vectors by E and I, then $\frac{E}{I}$ will be a vector; let it be the vector $Z = |Z| \,/\varphi$. To divide these vectors, divide the lengths and subtract the angles;

$$\text{then } |Z| = \frac{|E|}{|I|} = \frac{E_{max}}{I_{max}},$$

which is constant. Also, the angle φ will be equal to the angle between E and I, that is, the phase angle, which is also constant. Provided, then, that vectors are used throughout, the relationship $\frac{E}{I} = Z$ holds for AC, just as $\frac{E}{I} = R$ applies to DC. The vector Z is called the *impedance vector*, impedance being, in a sense, the AC equivalent of resistance, since it represents the total opposition of the circuit to current. Note that although E and I are rotating vectors, Z is fixed.

Since, in practical problems, alternating voltages and currents are measured by their RMS values, it is convenient to let the magnitudes of the vectors employed correspond to these RMS values.

Hence let $|E| = E_{RMS}$

and $|I| = I_{RMS}$

By so doing, the vector approach will yield an RMS answer. This convention will be used throughout this book. When no ambiguity is likely to arise between the vector and its modulus, the modulus sign may be omitted and the RMS voltages and currents represented by V and I.

The procedure for AC problems is first to find an expression for Z. The *vector* Z is of no use except as a means to an end; its modulus, however, gives the ratio between peak (or RMS) values $\frac{E_{max}}{I_{max}} = \frac{V_{RMS}}{I_{RMS}} = |Z|$ and its angle gives the phase angle by which the *voltage leads the current*. If the angle is negative the current is ahead of the voltage.

The value of $|Z|$ is the impedance of the circuit in ohms. Z, being a vector, can be expressed in the "notation" $R + jX$, then R is the resistive part of the impedance and X the reactive part, called the "reactance". The expression $\frac{1}{Z}$ is called the "admittance" and may be denoted by Y. It also is a vector.

The impedance Z of any circuit can be calculated provided that the impedance of each of the three fundamental components R, L and C is known; these three cases will now be dealt with.

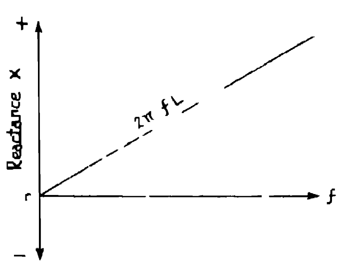


Fig. 154 — Reactance of an inductance

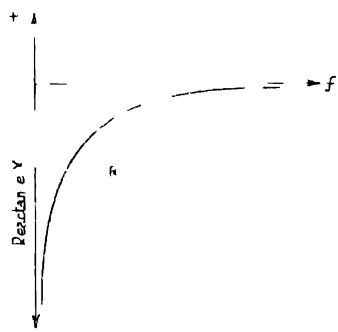


Fig. 155 — Reactance of a condenser

Resistance (see Fig. 148)

$$|Z| = \frac{V}{I} = R \qquad \text{angle } \phi = 0°,$$

$$Z = R\underline{/0°} = R \tag{10}$$

Hence the vector impedance of a resistance is a real number equal to its resistance and does not vary with frequency.

Inductance (see Fig. 150) —

$$|Z| = \frac{V_{max}}{I_{max}} = \omega L \qquad \text{angle } \phi = +90°$$

$$\therefore \quad Z = \omega L \underline{/90°} = j\omega L = j2\pi fL \tag{11}$$

Hence the vector impedance of an inductance is a pure imaginary quantity, whose magnitude increases with frequency. The reactance $X_L = \omega L = 2\pi fL$ and varies with frequency as in Fig 154.

Capacity (*see* Fig 152) —

$$|Z| = \frac{E_{max}}{I_{max}} = \frac{1}{\omega C}, \qquad \text{angle } \varphi = -90^\circ$$

$$\therefore \quad Z = \frac{1}{\omega C} \underline{/-90^\circ}$$

$$= \frac{-j}{\omega C} = \frac{1}{j\omega C} \tag{12}$$

These two expressions are equivalent. But *it will be found as a general rule that the first is the more useful if the condenser is in a series circuit and the second if it is in a parallel circuit.*

It will be seen that the vector impedance of a capacity is a pure imaginary quantity whose magnitude decreases with frequency. The reactance $X_C = \frac{-1}{\omega C} = \frac{1}{2\pi fC}$ (see Fig 155).

The three results just obtained are important as they form the basis of all AC theory. It is possible from them to calculate the impedance of any network, for when dealing with vectors the same methods can be employed as in DC. If, for instance, impedances are connected in series the total impedance is found by adding the *vector* impedances. It is most important to note that impedances cannot be added unless they are in vector form. Similarly if impedances are connected in parallel the total impedance may be found by adding the *vector* admittances.

Calculation of impedance of simple circuits

The simplest type of problem is the calculation of the impedance of a given circuit at some frequency. Two answers are generally required, the impedance in polar form and the phase angle between current and voltage, that is $|Z|$ and the angle φ. The angle φ is normally assumed to lie in either the first or the fourth quadrant. The steps in this type of problem are as follows:

(*a*) Draw the circuit inserting the impedances as vectors. If numerical values are given simplify each impedance as far as possible. It is usually best *not* to eliminate fractions.

(*b*) Find total *vector* impedance by applying the same methods as for DC problems. Simplify the answer if possible.

(*c*) Find $|Z|$. From this if the applied RMS voltage is given the RMS current may be found or *vice versa*, by remembering that $|Z| = \frac{E}{I} = \frac{E_{max}}{I_{max}}$.

(*d*) Find the angle φ. Hence the phase relationship between e and i is determined. Remember that if φ is negative the current is ahead of the voltage.

Two examples will illustrate this method.

Example 1 —A general problem without numerical values. An inductance and resistance are connected in series, find the *impedance at any frequency and the phase angle*

FIG. 156 Inductance in series with resistance

Step a —Consider Fig 156

Note that the impedance of the inductance has been written in vector form as $j\omega L$

Step b The two impedances as vectors are R and $j\omega L$. They are in series so the total impedance is found by adding —

$$Z = R + j\omega L$$

*Steps c and d**

$$\text{Hence} \quad |Z| = \sqrt{R^2 + \omega^2 L^2} \quad \text{Ans } (i)$$

$$\text{and} \quad \tan \varphi = \frac{\omega L}{R} \quad \text{Ans } (ii)$$

Since φ is positive the current lags on the voltage

The answers just obtained are general and could be applied to a particular case by substituting numerical values As, however, there are so many possible combinations of R L and C it is impossible to remember all the necessary formulae and it is better to work out each example from first principles

Example 2 – Find the current flowing in the circuit of Fig 157, and calculate its phase with respect to the voltage

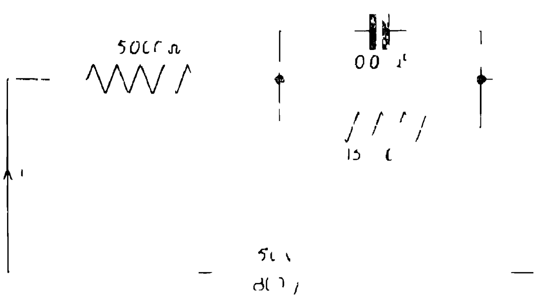


FIG. 157 – Circuit containing resistance in parallel with capacity in series with another resistance

Step a The impedance of the condenser must be written as a vector The two possible expressions are $-\frac{j}{\omega C}$ and $\frac{1}{j\omega C}$, and since this is a parallel circuit the second expression will be used

* See conversion from rectangular to polar vector notation, p 58.

here. Numerical values are now inserted, remembering that $f = 800$ gives $\omega = 5{,}000$. Then,

$$\frac{1}{j\omega C} = \frac{100 \times 10^6}{j \times 5 \times 10^3 \times 2} = \frac{10^8}{j \,.\, 10^4} = \frac{10^4}{j} = -j \,.\, 10{,}000$$

Step b.—To find the total impedance, the impedance of the parallel circuit must first be found. Remembering that :—

$$\frac{1}{Z_T} = \frac{1}{Z_1} + \frac{1}{Z_2}, \text{ so that } Z_T = \frac{1}{\dfrac{1}{Z_1} + \dfrac{1}{Z_2}},$$

this is given by :

$$Z_T = \frac{1}{\dfrac{1}{15{,}000} + \dfrac{j}{10^4}}$$

Bringing the bottom line over a common denominator :—

$$Z_T = \frac{1}{\dfrac{2 + j3}{30{,}000}} = \frac{3 \times 10^4}{2 + j3}$$

The total impedance Z is this in series with $5{,}000\,\Omega$

$$\therefore \qquad Z = 5 \times 10^3 + \frac{3 \times 10^4}{2 + j3}$$

Always remove common factors where possible ; in this case take out 5×10^3, giving

$$Z = 5 \times 10^3 \left[1 + \frac{3 \times 2}{2 + j3}\right]$$

Simplifying the expression in the bracket,

$$Z = 5 \times 10^3 \left[\frac{2 + j3 + 6}{2 + j3}\right]$$

$$\therefore \qquad Z = 5 \times 10^3 \frac{(8 + j3)}{(2 + j3)} \qquad . \qquad \text{(i)}$$

This is the answer in its simplest form. It is *not* usual to rationalise at this stage, for although it is easy in this case, it usually leads to much unnecessary additional work.

Step c.—Find $|Z|$. Z consists of a number of vectors multiplied together or divided ; to find its modulus the separate moduli must be multiplied or divided. The factor 5×10^3 of course remains unchanged, as can be seen by regarding it as a vector.

$$\text{Hence } |Z| = 5 \times 10^3 \times \frac{|8 + j3|}{|2 + j3|}$$

$$= 5 \times 10^3 \times \frac{\sqrt{64 + 9}}{\sqrt{4 + 9}}$$

$$\therefore \qquad |Z| = 5 \times 10^3\sqrt{\frac{73}{13}}$$

As the impedance itself is not required, this answer is not worked out at this stage. The current is calculated by remembering that $|Z| = \frac{E}{I}$

Now E = 50 volts,

$$\therefore \qquad I = \frac{50}{|Z|} = \frac{50}{5 \times 10^3}\sqrt{\frac{13}{73}} \text{ amps}$$

$$= 10\sqrt{\frac{13}{73}} \text{ mA}$$

$$= 4.2 \text{ mA} \quad \textit{Ans (i)}$$

Step d—Find the angle φ. Z consists of a number of vectors multiplied together or divided and its angle is found by adding or subtracting the individual angles.

Hence from (i) $\varphi = \tan^{-1}\left(\frac{3}{8}\right) - \tan^{-1}\left(\frac{3}{2}\right)$

$= \tan^{-1}(0.375) - \tan^{-1}(1.5)$

$= 20°\ 33' - 56°\ 18'$ (these values being chosen because $8 + j3$ and $2 + j3$ both lie in the first quadrant)

$= -35°\ 45'$ *Ans (ii)*

The negative sign shows that i leads e, as is to be expected in a capacitive circuit.

i is therefore a current of 4.2 mA leading the voltage by 35° 45′.

Calculation of frequency or component values in simple circuits

In some problems the impedance is given and either the frequency or one of the component values is required. Each problem must be treated on its merits, but two examples are given to illustrate typical methods.

Example 1 (see Fig. 158). Find the frequency.

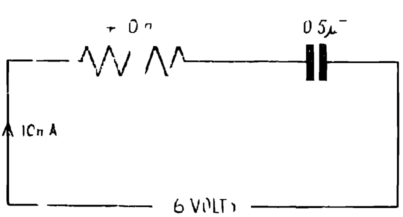


FIG. 158

As the voltage and current are given, the impedance can be found.—

$$|Z| = \frac{E}{I} = \frac{6 \times 1\,000}{10} = 600 \text{ ohms}$$

To obtain an equation for f, an expression is first found for the impedance of the resistance and condenser in series

Step a.—Consider Fig 159

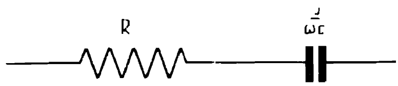


Fig 159

Step b.—
$$Z = R - \frac{j}{\omega C}$$

Step c.—
$$|Z| = \sqrt{R^2 + \frac{1}{\omega^2 C^2}}$$

Applying this expression to the example

$$600 = \sqrt{(400)^2 + \frac{(10^6)^2}{\omega^2 (0.5)^2}}$$

Squaring both sides:—

$$36 \times 10^4 = 16 \times 10^4 + \frac{10^{12}}{\omega^2} \times 4$$

$$\therefore \quad \frac{4 \times 10^{12}}{\omega^2} = 20 \times 10^4$$

$$\therefore \quad \omega^2 = \frac{4 \times 10^{12}}{20 \times 10^4} = 2 \times 10^7$$

$$\therefore \quad \omega = \sqrt{2 \times 10^7} = 10^3 \sqrt{20}$$

Thus
$$f = \frac{\omega}{2\pi} = \frac{10^3 \sqrt{20}}{2\pi}$$
$$= 712 \text{ c/s} \quad \textit{Ans}$$

Example 2 (*see* Fig 160)

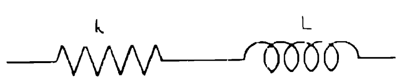


Fig 160

$|Z|$ is given as 500Ω at 800 c/s (i.e. $\omega = 5 \times 10^3$) (i)

$|Z|$ is given as 800Ω at 1 600 c/s (i.e. $\omega = 10^4$) (ii)

Find R and L

It is first necessary to calculate the impedance of a resistance and inductance in series

Step a.—Write the impedance of L as $j\omega L$

Step b.—The total impedance $Z = R + j\omega L$

Step c — $$|Z| = \sqrt{R^2 + \omega^2 L^2}$$

Apply this to the problem giving two equations

(i) becomes $$500 = \sqrt{R^2 + 25 \cdot 10^6 L^2} \quad \text{(iii)}$$

and (ii) becomes $$800 = \sqrt{R^2 + 100 \cdot 10^6 L^2} \quad \text{(iv)}$$

Squaring each side of these two equations

(iii) becomes $$25 \cdot 10^4 = R^2 + 25 \cdot 10^6 L^2 \quad \text{(v)}$$

and (iv) becomes $$64 \cdot 10^4 = R^2 + 100 \cdot 10^6 L^2 \quad \text{(vi)}$$

Subtract $$39 \cdot 10^4 = 75 \cdot 10^6 L^2 \quad \text{(vii)}$$

$$\therefore \quad L^2 = \frac{39 \cdot 10^4}{75 \cdot 10^6}$$

$$L = \frac{1}{10}\sqrt{\frac{39}{75}} \text{ Henries}$$

$$= 100\sqrt{\frac{39}{75}} \text{ mH}$$

$$= \frac{100}{5}\sqrt{\frac{39}{3}} \text{ mH}$$

$$= 20\sqrt{13} \text{ mH}$$

$$= 72 \text{ mH} \qquad \textit{Ans (a)}$$

Now find R

Multiplying equation (v) by 3

$$75 \cdot 10^4 = 3R^2 + 75 \cdot 10^6 L^2$$

Equation (vii)

$$39 \cdot 10^4 = 75 \cdot 10^6 L^2$$

Subtracting —

$$3R^2 = 36 \cdot 10^4$$

$$R^2 = 12 \cdot 10^4$$

$$R = 346 \text{ ohms} \qquad \textit{Ans (b)}$$

More complicated circuits

In more complicated circuits all currents and voltages will be written down as vectors. These vectors are in fact rotating but as only their *relative* positions are important it is permissible to consider them in any position in which the angle between them is fixed. The position is generally chosen so that one of the vectors becomes horizontal — it will then be used as a reference. An example illustrates this.

Example See Fig. 161

Find the current I

It will be noted that in the general case the same graphical symbol is used for all impedances irrespective of whether they are resistive, inductive, or capacitive.

In this example, the impedances have already been expressed

as vectors. To solve the problem, it will be most convenient to take the EMF of 6 volts as the reference vector; it thus becomes the vector $6\underline{/0}$. Let the total generator current be y, so the current in the central arm is $(y - x)$. x and y are both vectors, and x is required. There are two unknowns, so two equations must be found; these are obtained by applying Kirchhoff's law round the two small closed networks.

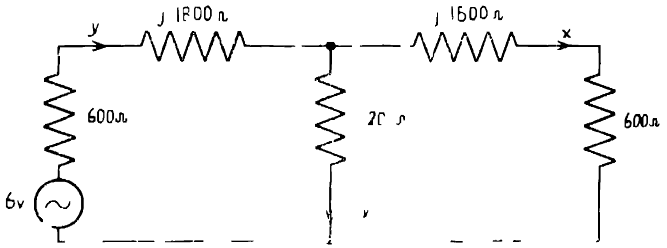


FIG. 11

Left-hand network:—

$$6 = y\,(600 + j1800 - j200) + j200x$$

$$\therefore\ 6 = y\,(600 + j1600) + j200x \qquad \text{(i)}$$

Right-hand network:—

$$0 = x\,(j1800 + 600 - j200) + j200y$$

$$0 = y\,j200 + x\,(600 + j1600)$$

$$0 = jy + x\,(3 + j8)$$

$$y = jx\,(3 + j8) \qquad \text{(ii)}$$

Substitute (ii) in (i) to eliminate y:—

$$6 = jx\,[(3 + j8)(600 + j1600) + 200]$$

$$= 200jx\,[1 + (3 + j8)^2] = 200jx\,[1 + 9 - 64 + j48]$$

$$\therefore\ x = \frac{6}{j200\,(j48 - 54)} = \frac{6}{200\,(-48 - j54)} = -\frac{1}{200\,(8 + j9)}$$

This gives x in vector form. Its modulus will give the current in amps.

$$|x| = \frac{1}{200\sqrt{64 + 81}}\ \text{amp} = \frac{1000}{200\sqrt{145}}\ \text{mA} = 0.42\ \text{mA} \quad \textit{Ans. (i)}$$

Its angle will be the angle by which x is ahead of the 6 volts, i.e.

$$180^\circ - \tan^{-1}\frac{9}{8} = 180^\circ - \tan^{-1} 1{\cdot}125 = 180^\circ - 48^\circ\,22'$$

$$= 131^\circ\,38' \quad \textit{Ans. (ii)}$$

(The "180°" is the angle corresponding to the " − 1 " in the numerator.)

POWER IN AC CIRCUITS

In DC, the power developed in any circuit is equal to the product of the voltage across it and the current through it. In AC, the *instantaneous power* is equal to the product of the instantaneous values of voltage and current, so that —

$$p = vi \qquad (13)$$

The instantaneous power is of little practical value; the important quantity is the "true power" which is defined as the *average value* of the power over a period of time and is measured in watts. The product of the RMS values $V \times I$ is called the *apparent power*; it might at first sight appear likely that this would give the true power, but it will be seen shortly that this is not so.

Calculation of true power in any circuit

Take the general case of a circuit when the voltage and current are out of phase by an angle φ so that

$$v = V_{max} \sin \omega t \qquad (14)$$

and

$$i = I_{max} \sin (\omega t + \varphi) \qquad (15)$$

(φ may be negative but this does not affect the answer.)

The instantaneous power is given by

$$p = V_{max} \sin \omega t \times I_{max} \sin (\omega t + \varphi)$$
$$= V_{max} I_{max} \sin \omega t \sin (\omega t + \varphi)$$

The true value of power P is the mean value of this; to calculate it the formula $2 \sin A \sin B = \cos (A - B) - \cos (A + B)$ is used.

This gives

$$p = \frac{V_{max} I_{max}}{2} [\cos \varphi - \cos (2\omega t + \varphi)]$$

Consider the expression inside the bracket. The first term is constant and its mean value over any period of time is $\cos \varphi$. The second term is a sinusoidal waveform and its mean value is zero.

Hence P = mean value of p

$$= \frac{V_{max} I_{max}}{2} \cos \varphi$$

i.e.

$$P = V \times I \cos \varphi \qquad (16)^*$$

since $V = \frac{V_{max}}{\sqrt{2}}$ and $I = \frac{I_{max}}{\sqrt{2}}$ from (2) and (3).

This shows that the true power depends not only on the values of the current and voltage but also on their relative phase. Cos φ is called the "power factor" of the load. Note that for a resistance the power factor is unity and the true power is equal to the apparent power. For a pure reactance (inductance or capacity) $\varphi = 90°$ and $\cos \varphi = 0$ and so *a pure reactance cannot absorb power*; this is a most important fact.

*If V and I refer to the corresponding rotating vectors then equation 16 becomes $P = |V| \times |I| \cos \varphi$ (see p. 199).

The equation just proved can be written as —

True power = Apparent power × power factor

If the impedance of the load can be written as —

$$Z = R + jX$$

then the power factor is given by —

$$\cos\varphi = \frac{R}{\sqrt{R^2 + X^2}} = \frac{R}{|Z|} \qquad (17)$$

Alternative forms of the expression can be calculated. Thus —

$$P = VI\cos\varphi = VI\frac{R}{|Z|}$$

$$= I^2|Z|\cos\varphi = I^2R = \frac{V^2R}{|Z|^2} = \frac{V^2}{|Z|}\cos\varphi \qquad (18)$$

Of these different forms $P = I^2R$ is perhaps the most useful. Note that the expression $P = \frac{V^2}{|Z|}$ is *not* the same as the formula $P = \frac{V^2}{R}$ in DC which applies only to a pure resistance.

As $I\cos\varphi$ is that component of I which is in phase with V, true power can be taken as the product of voltage and the in-phase component of current, or equally as the product of current and the in-phase component of voltage. This method is often useful in AC power questions.

A few examples will illustrate typical methods.

Example 1. See Fig. 162. Find the total power in watts.

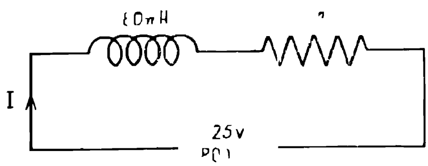


Fig. 162

The current I will be calculated and the formula I^2R applied. To calculate I, $|Z|$ must first be found.

$$\omega = 2\pi f = 2\pi \times 800 \approx 5000$$

The impedance of the inductance

$$X = j\omega L = j\,5000 \times \frac{80}{1000} = j400$$

$$|Z| = 100\sqrt{9 + 16} = 100\sqrt{25} = 500$$

$$I = \frac{V}{|Z|} = \frac{25}{500} = \frac{1}{20}$$

$$P = I^2R = \frac{1}{400} \times 300$$

$$= 0{\cdot}75 \text{ watts.}$$

The following is an alternative method :—

Taking the 25 volts as a reference vector, the current $I = \frac{25}{100\,(3 + j4)}$ The power is equal to 25 times the in-phase component of i, this can be found by rationalisation —

$$I = \frac{25}{100\,(3^2 + 4^2)}\,(3 - j4) = \frac{3}{100} - j\,\frac{4}{100}$$

The first term is the in phase component and the second, being multiplied by j, is 90° out of phase The power is therefore —

$$P = \frac{3}{100} \times 25 = 0.75 \text{ watt} \quad \textit{Ans}$$

Example 2 —*See* Fig 163

Find the total power.

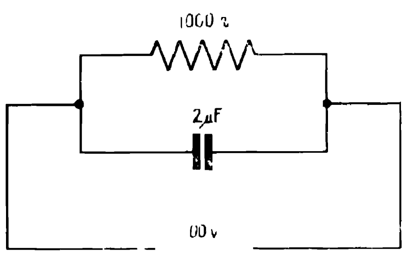


Fig 163

The answer should be obvious the condenser takes no power, and the power in the resistance is calculated from the voltage across it —

$$P = \frac{E^2}{R} = \frac{4 \times 10^4}{10^3} = 40 \text{ watt}$$

and this is the total power

Example 3 —A 250 volt 500 c s motor takes 1kW at a power factor of 0 8 lagging

(i) What current does it take ?

(ii) What condenser must be placed in parallel to bring the power factor to unity ?

(iii) What will then be the total supply current ?

(*a*) $1000 = EI \cos\varphi = 250 \times I \times 0.8$

$I = 5$ amps *Ans* (*i*)

The in-phase component of this current is

$$I \cos\varphi = 5 \times 0.8 = 4 \text{ amps}$$

and the 90° out-of-phase component is

$$I \sin\varphi = I\sqrt{1 - \cos^2\varphi} = 5\sqrt{1 - 0{\cdot}64} = 3 \text{ amps}$$

(*b*) To bring the power factor to unity, a condenser must be added that will take a current of 3 amps, 90° *ahead* of the voltage.

It must therefore have a reactance of $\frac{250}{3}$ ohms.

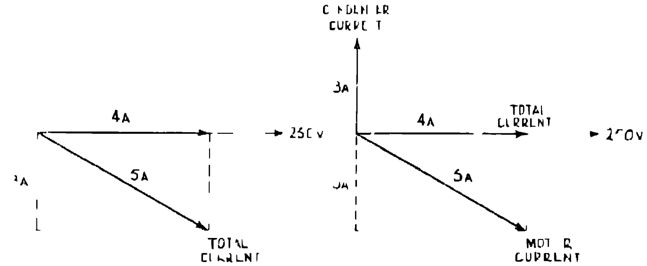
$$\therefore \quad \frac{1}{\omega C} = \frac{10^6}{2\pi \times 500 \times C} = \frac{250}{3} \quad \text{(with } C \text{ in } \mu\text{F)}$$

$$\therefore \quad C = \frac{3 \times 10^6}{250 \times 1000\pi} = \frac{12}{\pi}$$

$$= 3{\cdot}82\ \mu\text{F} \quad \textit{Ans (ii)}$$

(c) The total line current will then be merely the in-phase component that is, 4 amps *Ans (iii)*

Note that by adding a condenser to increase the power factor, the total current has been decreased Fig 164 shows the vector diagrams with and without the condenser.



(a) WITHOUT CONDENSER (b) WITH CONDENSER

FIG 164 —Effect of power factor correcting condenser

With large electrical installations, particular attention has to be paid to the power factor, the nearer it approaches unity, the smaller is the total current, and the smaller the loss in the distribution lines.

RESONANCE

In AC circuits (apart from those consisting of pure resistances), the current and voltage are usually out of phase Under certain conditions, however, they may be in phase, and the circuit behaves as a resistance This phenomenon is known as resonance For a given circuit it would normally occur at only a finite number of frequencies usually one, in the simplest cases. It is possible, however, to find circuits that resonate at all frequencies

A two-terminal network containing reactance is said to resonate when the voltage across it and current through it are in phase

This is the general definition of resonance, resonance should not be confused with the condition for maximum or minimum impedance, though this often occurs at or near resonance.

A few particular circuits will now be considered. It is normally required to calculate the frequency at which a given circuit will

resonate, though it is also often necessary to calculate the value of one or more of the components that will make a circuit resonate at a given frequency. The impedance at resonance is important and will be found in each case. The resonant frequency is denoted by f_0, and the corresponding angular velocity by ω_0.

General rules for finding the condition for resonance

To find the condition for resonance, it is necessary simply to write down the impedance Z, and to state the condition that this shall be resistive (*i.e.*, real). This can be done in a variety of ways, all of which are worth bearing in mind.

If Z is in the form $[r,\theta]$, the resonant condition is $\theta = 0$.

If Z is in the form $R + jX$, the resonant condition is $X = 0$.

If $Y = Z^{-1}$ is in the form $G + jB$, the resonant condition is $B = 0$.

If Z is in the form $\dfrac{A + jB}{C + jD}$, the resonant condition is $\dfrac{B}{A} = \dfrac{D}{C}$.

(This last condition is equivalent to saying that the numerator and denominator have equal angles and hence that the angle of Z is 0.) From the condition for resonance, the resonant frequency can be calculated.

Series resonance

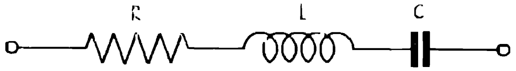


FIG. 165.—Series resonant circuit.

In Fig. 165, the impedance is :—

$$Z = R + j\omega L - \frac{j}{\omega C}$$

$$= R + j\left(\omega L - \frac{1}{\omega C}\right) \tag{19}$$

The condition for resonance is :—

$$\omega L - \frac{1}{\omega C} = 0$$

$$\therefore \qquad \omega^2 = \frac{1}{LC}$$

Thus $$\omega = \omega_0 = \frac{1}{\sqrt{LC}}, \text{ or } f = f_0 = \frac{1}{2\pi\sqrt{LC}} \tag{20}$$

From equation 19, it can be seen that the impedance at resonance (when $X = 0$) is equal to R.

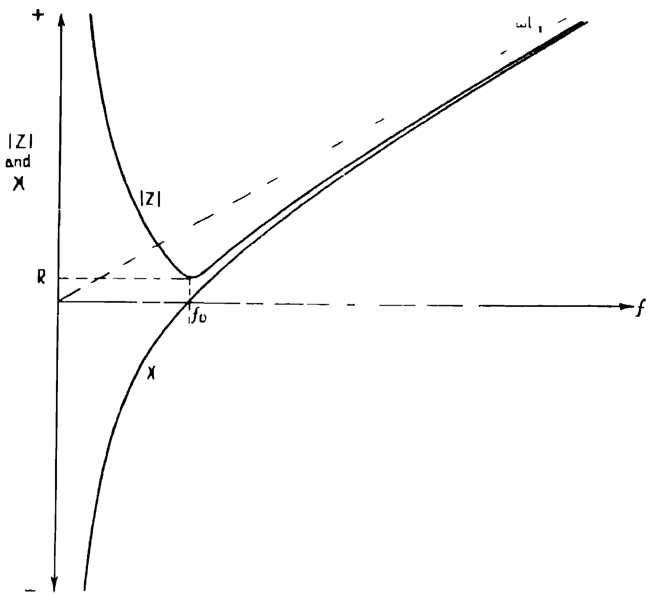


FIG. 166.—Variation of impedance and reactance with frequency for a series circuit.

Note that at any other frequency, the magnitude of the impedance is :—

$$|Z| = \sqrt{R^2 + (\omega L - \frac{1}{\omega C})^2} \tag{21}$$

which is always greater than R. Hence the impedance is a *minimum* at resonance. Fig. 166 shows how the impedance and reactance vary with frequency.

Voltages in series resonance circuits

Let a voltage E be applied to the above circuit at its resonant frequency. Consider the voltage across the condenser (E_C) or inductance (E_L); this is calculated from the total current.

$$I = \frac{E}{R}, \text{ and } E_L = I \times \omega_0 L$$

$$\therefore \qquad E_L = E \times \frac{\omega_0 L}{R}$$

Denote $\frac{\omega_0 L}{R}$ by the letter Q.

Then $\qquad E_L = Q \times E$

Similarly, $E_C = \dfrac{I}{\omega_0 C} = I\omega_0 L$, since, at resonance, $\dfrac{1}{\omega_0 C} = \omega_0 L$

$$\therefore \qquad E_C = E_L = Q \times E \qquad (22)$$

$Q = \dfrac{\omega_0 L}{R}$ may often be very large; if R is the resistance of the inductance only, Q may have values up to several hundred. Hence the voltage across the inductance or condenser may be several hundred times the voltage across the whole circuit. This important property of series resonant circuits is of wide application in communication engineering, as it provides a simple means of discriminating between different frequencies. If a constant voltage at

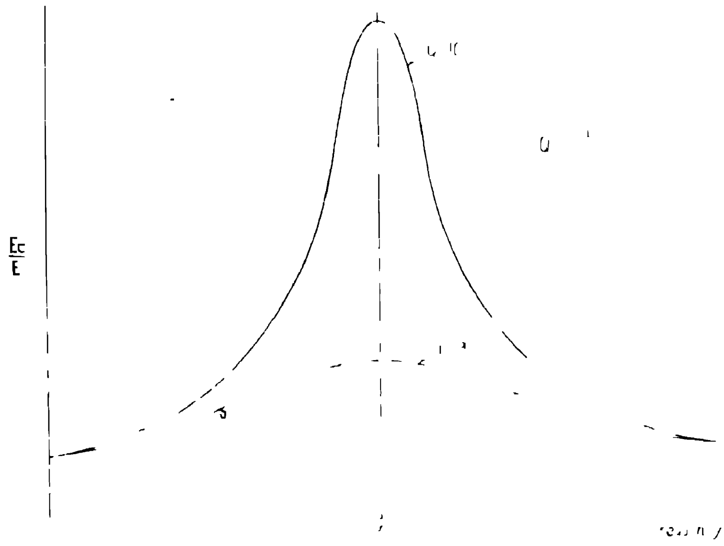


FIG 167 —Variation with frequency of voltage across condenser of Fig 165 for constant applied voltage

varying frequency is applied to the circuit, the voltage across, say, the condenser will reach a maximum just below the resonant frequency, but if Q is large the difference between the two frequencies is small. Fig. 167 shows the variation of $\dfrac{E_C}{E}$ with frequency.

The ratio $$\frac{E_C}{E} = \frac{|\text{condenser impedance}|}{|\text{total impedance}|}$$

$$= \frac{1}{\omega C\sqrt{R^2 + \left(\omega L - \dfrac{1}{\omega C}\right)^2}}$$

Let this equal n. The maximum value of n^2 will occur at the same frequency as that of n.

$$n^2 = \frac{1}{\omega^2 C^2\left(R^2 + \omega^2 L^2 - \dfrac{2L}{C} + \dfrac{1}{\omega^2 C^2}\right)}$$

$$= \frac{1}{\omega^2 C^2\left(R^2 - \dfrac{2L}{C}\right) + \omega^4 L^2 C^2 + 1}$$

To find the frequency that makes n^2 a maximum, differentiate (for simplicity, with respect to ω^2) and put $\dfrac{d\,(n^2)}{d\,(\omega^2)} = 0$

$$\frac{d\,(n^2)}{d\,(\omega^2)} = -\frac{\left[C^2\left(R^2 - \dfrac{2L}{C}\right) + 2\,\omega^2 L^2 C^2\right]}{\left[\omega^2 C^2\left(R^2 - \dfrac{2L}{C}\right) + \omega^4 L^2 C^2 + 1\right]^2}$$

$$\therefore \quad 2\,\omega^2 L^2 C^2 = -\,C^2\left(R^2 - \frac{2L}{C}\right)$$

$$\therefore \quad \omega^2 L^2 = \frac{L}{C} - \frac{R^2}{2}$$

$$\therefore \quad \omega^2 = \frac{1}{LC} - \frac{R^2}{2L^2}$$

$$= \omega_0^2 - \frac{R^2}{2L^2} = \omega_0^2\left(1 - \frac{R^2}{2\,\omega_0^2 L^2}\right)$$

$$= \omega_0^2\left(1 - \frac{1}{2Q^2}\right) \text{ since } Q = \frac{\omega_0 L}{R}$$

$$\therefore \quad \omega = \omega_0\sqrt{1 - \frac{1}{2Q^2}} \quad \therefore \quad f = f_0\sqrt{1 - \frac{1}{2Q^2}} \qquad (23)$$

f is therefore less than f_0 but if Q is large ($>$ 10, say), the term $\dfrac{1}{2Q^2} \simeq 0$ and $f \simeq f_0$

Parallel resonance (anti-resonance)

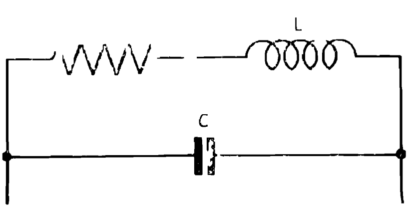


FIG 168 Parallel circuit containing R, L and C

The case of a pure inductance in parallel with a condenser is not considered as in practice an inductance always possesses resistance. The circuit to be dealt with is shown in Fig 168

The impedance is given by :—

$$\frac{1}{Z}=\frac{1}{R+j\omega L}+j\omega C$$

The simplest way of determining the condition for resonanc is by rationalisation.

$$\frac{1}{Z}=\frac{R-j\omega L}{(R+j\omega L)(R-j\omega L)}+j\omega C$$

$$\therefore \quad Y=Z^{-1}=\frac{R}{R^2+\omega^2L^2}+j\omega\left[C-\frac{L}{R^2+\omega^2L^2}\right] \qquad (24)$$

The condition for resonance is that the "j" term is equal to zero.

This will occur when $C=\dfrac{L}{R^2+\omega^2L^2}$ (25)

$$R^2+\omega^2L^2=\frac{L}{C}$$

$$\omega^2L^2=\frac{L}{C}-R^2$$

$$\omega^2=\frac{1}{LC}-\frac{R^2}{L^2}$$

$$\omega=\omega_0=\sqrt{\frac{1}{LC}-\frac{R^2}{L^2}}$$

and
$$f_0=\frac{1}{2\pi}\sqrt{\frac{1}{LC}-\frac{R^2}{L^2}} \qquad (26)$$

The impedance at resonance is given by equation 24 when the "j" term is put equal to zero.

$$\therefore \quad \frac{1}{Z}=\frac{R}{R^2+\omega_0^2L^2}$$

$$=\frac{R}{\frac{L}{C}} \text{ (from equation 25)}$$

$$\therefore \quad Z=\frac{L}{CR} \text{ at resonance} \qquad (27)$$

Note that if R is very small, Z will be very large, tending to infinity as R approaches zero.

If R is small, there is a useful approximation to this formula or Z.

For
$$\frac{1}{Z}=\frac{R}{R^2+\omega_0^2L^2}, \text{ as above,}$$

$$\therefore \quad Z=R+\frac{\omega_0^2L^2}{R}$$

Neglecting the first term :—

$$Z \simeq \frac{\omega_0^2 L^2}{R} = \frac{\omega_0 L}{R} \cdot \omega_0 L$$

$$= Q \cdot \omega_0 L \qquad (28)$$

where $Q = \dfrac{\omega_0 L}{R}$, as for series resonance. If Q is large, it can be seen that the impedance of the circuit at resonance is much greater than the impedance of the inductance.

It can be shown that, if Q is large, $|Z|$ is a maximum at resonance. In theory, the maximum value occurs at a frequency

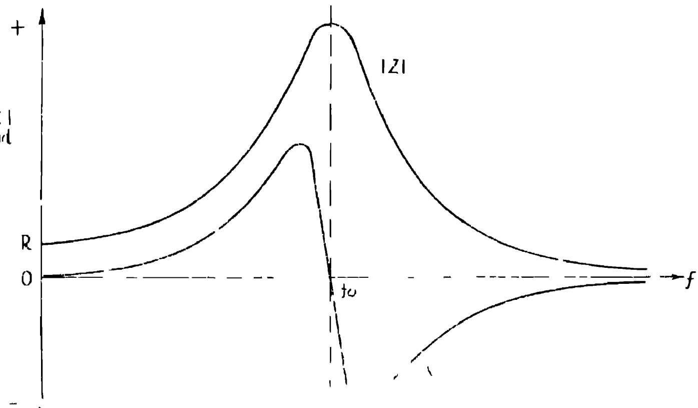


FIG. 169 Variation of impedance and reactance for a parallel circuit.

just above the resonant frequency, but the difference between the two frequencies is small. Fig 169 shows how the impedance and reactance vary with frequency.

Currents in parallel resonant circuit

Let the total current through the circuit be I at the resonant frequency; consider the current through the condenser (I_C) or inductance (I_L).

This can be found from the voltage across the circuit, which is equal to the product of I and the impedance at resonance,

i.e., $$E = I \cdot \frac{L}{CR}$$

$$I_C = E \cdot \omega_0 C = \frac{IL}{CR} \times \omega_0 C = I \times \frac{\omega_0 L}{R}$$

$$= Q \cdot I \qquad (29)$$

If Q is large, it can be shown that I_L is also roughly equal to $Q \times I$.

Hence in a parallel resonant circuit, the current through either arm is much greater than the total current.

A few examples will illustrate the application of these formulae.

Example 1.—Find the series and parallel resonant frequencies of a condenser of 0·005 μF and an inductance of 100 mH whose resistance is 500 ohms. Find also the impedance at resonance of the parallel circuit.

Here $\qquad L = 0 \cdot 1,\ C = 5 \times 10^{-9}$, and $R = 500$

For series resonance,

$$f_0 = \frac{1}{2\pi\sqrt{LC}} = \frac{1}{2\pi\sqrt{10^{-1} \times 5 \times 10^{-9}}} = \frac{10^5}{2\pi\sqrt{5}}$$
$$= 7119 \text{ c/s} \quad \textit{Ans.}\ (i)$$

For parallel resonance,

$$f_0 = \frac{1}{2\pi}\sqrt{\frac{1}{LC} - \frac{R^2}{L^2}} = \frac{1}{2\pi}\sqrt{\frac{10^{10}}{5} - 25 \times 10^4 \times 10^2}$$
$$= \frac{1}{2\pi}\sqrt{10^6\,(2000 - 25)} = \frac{10^3}{2\pi}\sqrt{1975}$$
$$= 7073 \text{ c/s} \quad \textit{Ans.}\ (ii)$$

The impedance at resonance $= \dfrac{L}{CR}$

$$= \frac{10^9}{10 \times 5 \times 500} = \frac{10^6}{25} = 40{,}000 \text{ ohms.} \quad \textit{Ans.}\ (iii)$$

Note that although R is not small (Q is about 10), the difference between the two resonant frequencies is small.

Example 2.—A coil of inductance 1H and resistance 300 ohms resonates with a series condenser at 500 c/s. What voltage will appear across the condenser if 10 volts is applied to the circuit (*a*) at 500 c/s? (*b*) at 1000 c/s?

The value of the condenser is not required, and need not be found for this problem.

At 500 c/s, the reactance of the inductance will be $j2\pi fL = j1000\pi$. The condenser must therefore have a reactance of $-j1000\pi$ at 500 c/s.

$$Q = \frac{2\pi fL}{R} = \frac{1000\pi}{300}$$

If 10 volts is applied to the whole circuit, the voltage across C will be :—

$$10 \times Q = \frac{1000\pi}{30} = 104 \cdot 5 \text{ volts.} \quad \textit{Ans.}\ (i)$$

At 1000 c/s, the reactance of L will have doubled, $= j2000\pi$ and the reactance of C will have halved, $= -j500\pi$

$$\therefore \text{ total impedance} = 300 + j2000\pi - j500\pi = 300 + j1500\pi$$
$$= 300\,(1 + j5\pi)$$

$$\therefore \qquad |Z| = 300\sqrt{1 + 25\pi^2}$$

The voltage across the condenser will be equal to the product of total current and impedance of the condenser, i e, $\frac{10}{|Z|} \times 500\pi$

$$= \frac{10 \times 500\pi}{300\sqrt{1 + 25\pi^2}} = \frac{50\pi}{3\sqrt{1 + 246}} = \frac{50\pi}{3\sqrt{247}} = 3\cdot33 \text{ volts } \textit{Ans} \text{ (ii)}$$

This example gives a good illustration of the "magnification" effect of a series circuit

Selectivity of resonant circuits

Series and parallel resonant circuits can be used to pick out signals at their resonant frequency Consider Fig 170 -

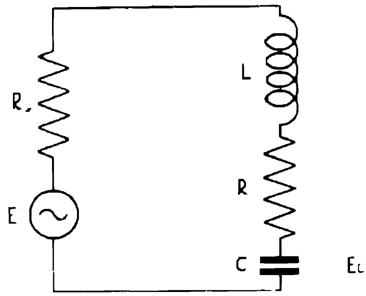


Fig 170 Generator connected to series resonant circuit

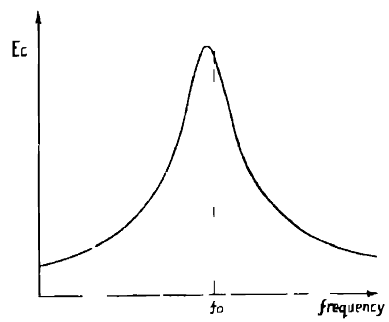


Fig 171 Variation with frequency of voltage across condenser of Fig 165 for constant applied voltage when generator impedance is low

Suppose that a generator of constant voltage E and internal impedance R_g is connected to a series circuit and that the frequency of E is varied Consider the voltage E_c across C As the frequency is varied, the total impedance of L C and R will vary, if R_g is always small compared with this impedance the voltage across L C and R will be approximately equal to E At resonance E_c will be approximately equal to Q E The variation of E_c with f is shown in Fig 171 It can be seen that it reaches a sharp maximum just below resonance the circuit is "selective"

Consider the same circuit but with R_g large If the impedance of L, C and R is always very small compared with R_g, the current in the circuit will be approximately equal to $\frac{E}{R_g}$ at all frequencies. In this case, the voltage across C will drop steadily with frequency, as shown in Fig 172.

This circuit is clearly not selective It will thus be seen that a series circuit will operate satisfactorily as a selective circuit

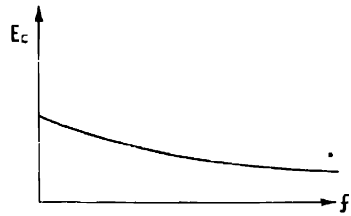


FIG. 172 Variation with frequency of voltage across condenser of Fig. 165 for constant applied voltage, when generator impedance is high

only if supplied from a generator having a low internal impedance.

Fig. 173 shows a parallel circuit. Consider the voltage E' across this circuit when it is connected to the generator. Suppose that E is kept constant and the frequency is varied. If R_G is small at all frequencies compared with the impedance of the parallel circuit, then E' will be approximately equal to E, and the circuit will not be selective. If, however, R_G is always large compared with the impedance of the circuit, then the current flowing will always be approximately equal to $\frac{E}{R_G}$. In this case, E' will be proportional to the impedance of the parallel circuit, and will therefore be a maximum at resonance as shown in Fig. 174.

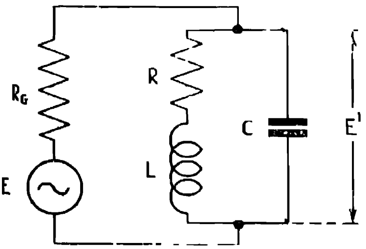


FIG. 173 Generator connected to parallel resonant circuit

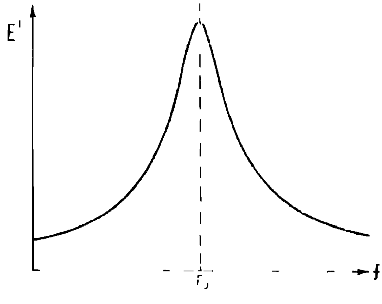


FIG. 174 - Voltage across parallel resonant circuit.

Hence the parallel circuit is selective only if supplied from a generator having a high internal impedance.

Parallel circuit resonant at all frequencies

An example of a circuit that resonates at all frequencies will now be given. It is of wide application in the design of constant-impedance reactive networks. Consider the circuit shown in Fig. 175.

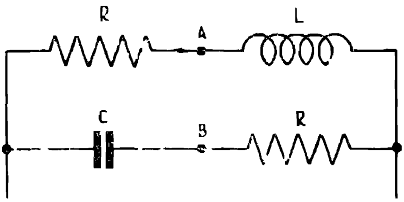


FIG. 175.—Parallel circuit with equal resistance in both arms.

Note that the two resistances are equal. The condition of resonance will first be obtained.

$$\frac{1}{Z} = \frac{1}{R + j\omega L} + \frac{1}{R - \frac{j}{\omega C}} = \frac{R - \frac{j}{\omega C} + R + j\omega L}{(R + j\omega L)\left(R - \frac{j}{\omega C}\right)}$$

$$\therefore \quad \frac{1}{Z} = \frac{2R + j\left(\omega L - \frac{1}{\omega C}\right)}{R^2 + \frac{L}{C} + jR\left(\omega L - \frac{1}{\omega C}\right)} \qquad (30)$$

The condition for resonance is that the numerator and denominator should have the same angle,

$$\text{i.e.} \qquad \frac{\left(\omega L - \frac{1}{\omega C}\right)}{2R} = \frac{R\left(\omega L - \frac{1}{\omega C}\right)}{R^2 + \frac{L}{C}}$$

Cross-multiplying—

$$\left(\omega L - \frac{1}{\omega C}\right)\left(R^2 + \frac{L}{C}\right) = 2R^2\left(\omega L - \frac{1}{\omega C}\right)$$

$$\therefore \quad \left(\omega L - \frac{1}{\omega C}\right)\left(R^2 - \frac{L}{C}\right) = 0$$

This is satisfied if $\omega L = \dfrac{1}{\omega C}$ (31)

or $R^2 = \dfrac{L}{C}$ (32)

Consider equation 31. This gives the normal frequency of resonance. At this frequency the impedance, from equation 30, will be $\dfrac{R^2 + \frac{L}{C}}{2R} = \dfrac{R}{2} + \dfrac{L}{2CR}$. This result is of no particular interest.

Consider equation 32. This is a relationship between the components of the circuit that would not normally be satisfied. It does not, however, involve frequency, and hence if it were

satisfied the circuit would be resonant at all frequencies; for if $R^2 = \frac{L}{C}$, equation 30 can be written as:—

$$Z = \frac{R^2 + R^2 + Rj\left(\omega L - \frac{1}{\omega C}\right)}{2R + j\left(\omega L - \frac{1}{\omega C}\right)} = \frac{R\left[2R + j\left(\omega L - \frac{1}{\omega C}\right)\right]}{2R + j\left(\omega L - \frac{1}{\omega C}\right)}$$

$$= R \qquad (33)$$

Hence if $R^2 = \frac{L}{C}$, the impedance of this circuit is equal to R at all frequencies. It can now be seen why this circuit is so important —for although it contains reactances, its impedance is constant and resistive at all frequencies.

It is worth noting that $R^2 = \frac{L}{C}$ is the condition that the bridge formed by the four components should be balanced; hence if this is true there will be no voltage across AB (*see* Fig. 175, p. 221).

The same results hold if any two impedances Z_1 and Z_2 are substituted for L and C provided that $Z_1 Z_2 = R^2$. Impedances that satisfy this condition are known as "inverse impedances with respect to R".

Reactance sketches

Reactance sketches may be employed to determine the resonant frequency of a network. The impedance (Z) of any circuit may be written in the form:—

$$Z = R + jX \qquad (34)$$

where R is the resistance and X the reactance.

The admittance (Y) of the circuit is defined as the reciprocal of the impedance and is also complex,

i.e.
$$Y = \frac{1}{Z} = G + jB \qquad (35)$$

where G is the "conductance" and B the "susceptance" of the circuit. (The susceptance is sometimes represented by S instead of B.)

From equations 34 and 35:—

$$G + jB = \frac{1}{R + jX} = \frac{R - jX}{R^2 + X^2}$$

$$\therefore \quad G = \frac{R}{R^2 + X^2} \qquad (36)$$

and
$$B = \frac{-X}{R^2 + X^2} \qquad (37)$$

Also
$$R + jX = \frac{1}{G + jB} = \frac{G - jB}{G^2 + B^2}$$

$$\therefore \qquad R = \frac{G}{G^2 + B^2} \tag{38}$$

and

$$X = \frac{-B}{G^2 + B^2} \tag{39}$$

In the case of series resonance, the condition for resonance was taken as the condition for zero reactance, i.e. $X = 0$, and this condition gives series resonance whether resistance is present or not. From equation 37 the susceptance at resonance is zero unless the resistance of the series circuit is also zero.

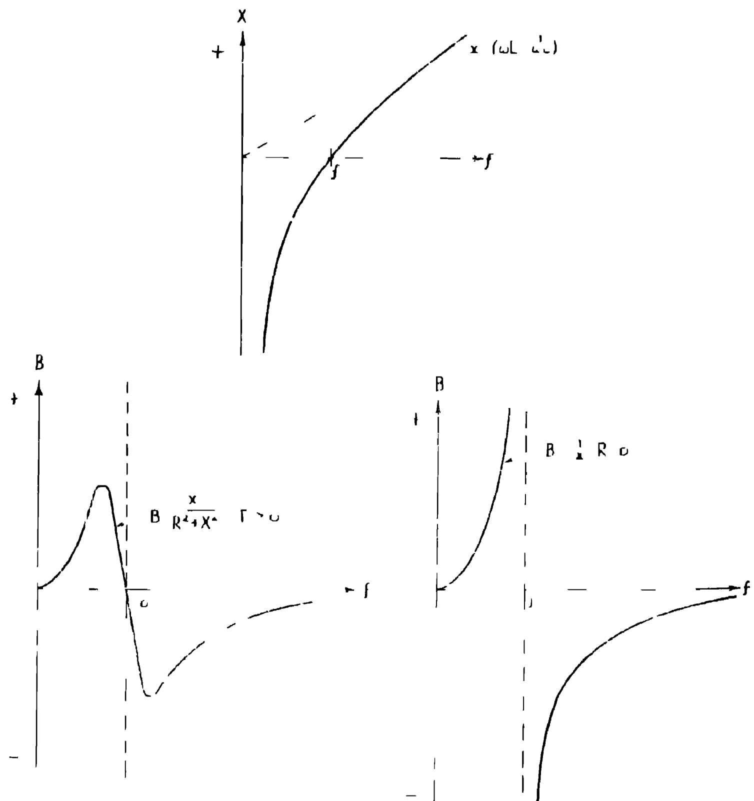


Fig. 176 Variation of reactance and susceptance with frequency for a series circuit

In the case of parallel resonance, the condition for resonance was found by equating to zero the imaginary part of $\frac{1}{Z}$, that is, the condition is one of zero susceptance. This condition $B = 0$ gives parallel resonance whether resistance is present or not.

The general shape of these curves (Figs. 176 and 177) should be

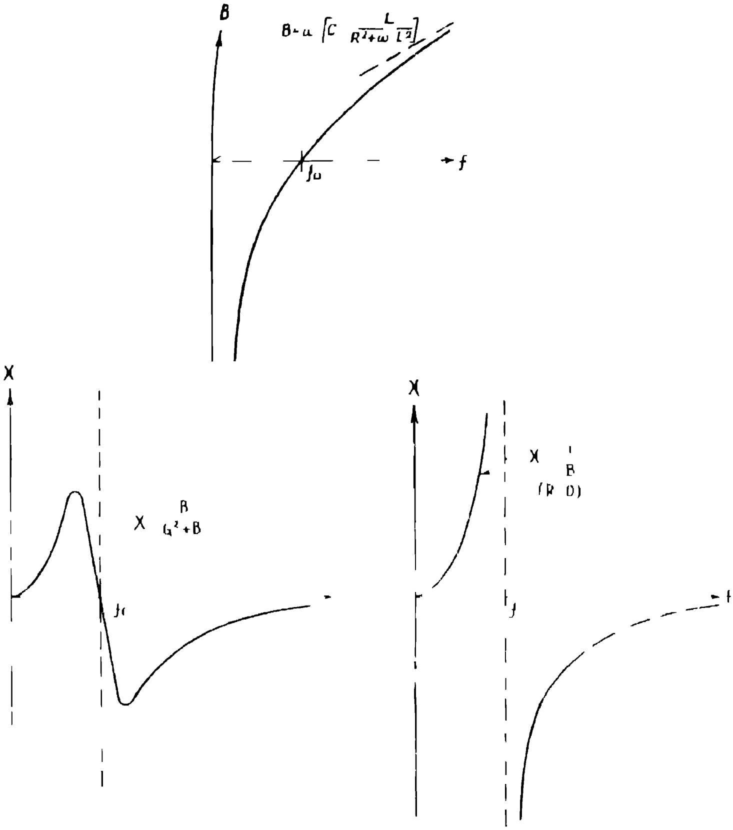


FIG. 177—Variation of susceptance and reactance with frequency for a parallel circuit

memorised as they are extremely useful in exploring the behaviour of filters and other complex networks. As an example of the use to which these curves can be put consider a circuit that has more than one resonant frequency.

Circuit having two resonant frequencies

Consider the circuit in Fig. 178. If this circuit is treated by the normal method:—

$$Z = \frac{\dfrac{-j}{\omega C}\left\{R + j\left(\omega L - \dfrac{1}{\omega C}\right)\right\}}{R + j\left(\omega L - \dfrac{2}{\omega C}\right)}$$

$$= \frac{\frac{-j}{\omega C}[\omega CR + j(\omega^2LC - 1)]}{\omega CR + j(\omega^2LC - 2)}$$

$$= \frac{\frac{-j}{\omega C}[\omega CR + j(\omega^2LC - 1)][\omega CR - j(\omega^2LC - 2)]}{\omega^2C^2R^2 + (\omega^2LC - 2)^2}$$

The reactance is given by :—

$$X = \frac{\frac{-1}{\omega C}[\omega^2C^2R^2 + (\omega^2LC - 1)(\omega^2LC - 2)]}{\omega^2C^2R^2 + (\omega^2LC - 2)^2} \qquad (40)$$

This is zero if :—

$$\omega^4L^2C^2 - \omega^2(3LC - C^2R^2) + 2 = 0$$

This has two positive roots

$$\omega = \sqrt{\frac{(3L - CR^2) \pm \sqrt{(3L - CR^2)^2 - 8L^2}}{2L^2C}} \qquad (41)$$

Thus the circuit has two resonant frequencies at both of which the reactance is zero. It can be shown that where a circuit has

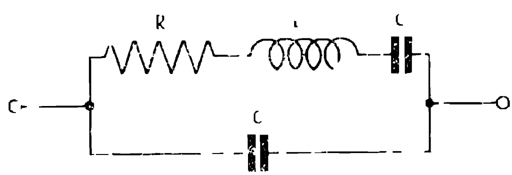


FIG. 178.—Simple circuit having two resonant frequencies.

more than one resonant frequency series and parallel resonances must occur alternately. This will not be proved but a method for exploring the resonances by means of reactance sketches will be given.

Consider the circuit of Fig. 178 but suppose that the resistance $R = 0$.

Fig. 179 shows how the reactance frequency sketch is built up; (*a*) and (*b*) show the susceptance frequency sketches for the two arms of the circuit and (*c*) shows the susceptance sketch of the whole circuit being the sum of the susceptances of the two parallel arms. Finally (*d*) shows the reactance frequency sketch obtained from (*c*) by the relationship $X = -\frac{1}{B}$ which holds for pure reactances. The frequency f_1 for which the reactance is zero is a series resonance; the frequency f_2, which gives infinite reactance, is a parallel resonance. The presence of a small resistance in the circuit will in general alter the *value* of these resonant frequencies, but it will not alter the *order* in which they occur.

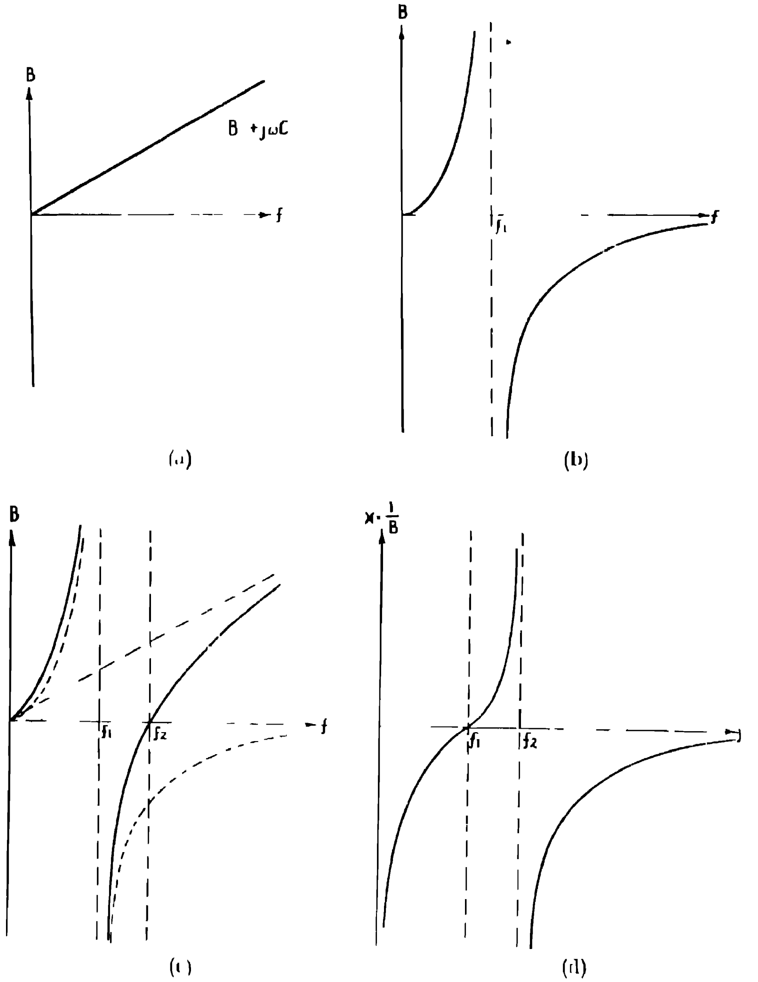


FIG. 179.—Method for obtaining the reactance frequency curves of the circuit of Fig. 178 (assuming $R = 0$)

Thus the lower frequency :—

$$f_1 = \frac{1}{2\pi}\sqrt{\frac{(3L - CR^2) - \sqrt{(3L - CR^2)^2 - 8L^2}}{2L^2C}} \tag{42}$$

gives series resonance, and the upper frequency :—

$$f_2 = \frac{1}{2\pi}\sqrt{\frac{(3L - CR^2) + \sqrt{(3L - CR^2)^2 - 8L^2}}{2L^2C}} \tag{43}$$

gives parallel resonance.

Summary of conditions for resonance, for maximum and minimum impedance, and for maximum voltages and currents

$\omega_0 = 2\pi \times$ series resonant frequency $= \dfrac{1}{\sqrt{L_0 C_0}}$

L_0 = value of inductance (L) to give series resonant condition

C_0 = value of capacity (C) to give series resonant condition.

$Q = \dfrac{\omega_0 L_0}{R}$

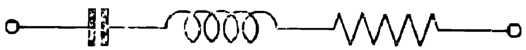


FIG. 180.—Series resonant circuit.

1. *Series circuit, with R assumed constant* (Q varies with ω).—

(*a*) Condition for resonance :—

$$\omega^2 LC = 1 \tag{44}$$

(*b*) Condition for minimum impedance on varying L, C or ω :—

$$\omega^2 LC = 1 \tag{45}$$

(*c*) Condition for maximum voltage across L, with constant applied voltage :—

(i) If ω be varied :

$$\omega^2 = \frac{1}{LC\left(1 + \dfrac{R^2C}{2L}\right)} = \omega_0^2 \cdot \left(1 + \frac{1}{2Q^2}\right)^{-1} \tag{46}$$

(ii) If L be varied :

$$L = \frac{1}{\omega^2 C} + CR^2 = L_0\left(1 + \frac{1}{Q^2}\right) \tag{47}$$

(iii) If C be varied :

$$C = \frac{1}{\omega^2 L} = C_0 \tag{48*}$$

(*d*) Condition for maximum voltage across C, with constant applied voltage :—

(i) If ω be varied :

$$\omega^2 = \frac{1}{LC} - \frac{R^2}{2L^2} = \omega_0^2 \cdot \left(1 - \frac{1}{2Q^2}\right) \tag{49}$$

(ii) If L be varied :

$$L = \frac{1}{\omega^2 C} = L_0 \tag{50*}$$

* *i.e.*, as for resonance.

(iii) If C be varied

$$C = \frac{L}{R^2 + \omega^2 L^2} = C_0\left(1 + \frac{1}{Q^2}\right)^{-1} \tag{51}$$

2 *Series circuit, with Q assumed constant* (R varies with ω) —

(*a*) Condition for resonance —

$$\omega^2 LC = 1 \tag{52}$$

(*b*) Condition for minimum impedance

(i) If ω be varied

$$\omega = \frac{1}{\sqrt{LC}}\left(1 + \frac{1}{Q^2}\right) = \omega_0\left(1 + \frac{1}{Q^2}\right) \tag{53}$$

(ii) If L be varied

$$L = \frac{1}{\omega^2 C}\left(1 + \frac{1}{Q^2}\right)^{-1} = L_0\left(1 + \frac{1}{Q^2}\right)^{-1} \tag{54}$$

(iii) If C be varied

$$C = \frac{1}{\omega^2 L} = C_0 \tag{55)*}$$

(*c*) Condition for maximum voltage across L with constant applied voltage

(i) If ω be varied

$$\omega^2 = \frac{1}{LC} = \omega_0 \tag{56)*}$$

(ii) If L be varied

$$L = \frac{1}{\omega^2 C} = L_0 \tag{57)*}$$

(iii) If C be varied

$$C = \frac{1}{\omega^2 L} = C_0 \tag{58)*}$$

(*d*) Condition for maximum voltage across C with constant applied voltage

(i) If ω be varied

$$\omega = \frac{1}{\sqrt{LC}}\left(1 + \frac{1}{Q^2}\right)^{-1} = \omega_0\left(1 + \frac{1}{Q^2}\right)^{-1} \tag{59}$$

(ii) If L be varied

$$L = \frac{1}{\omega^2 C}\left(1 + \frac{1}{Q^2}\right)^{-1} = L_0\left(1 + \frac{1}{Q^2}\right)^{-1} \tag{60}$$

(iii) If C be varied

$$C = \frac{1}{\omega^2 L}\left(1 + \frac{1}{Q^2}\right)^{-1} = C_0\left(1 + \frac{1}{Q^2}\right)^{-1} \tag{61}$$

* *i e* as for resonance

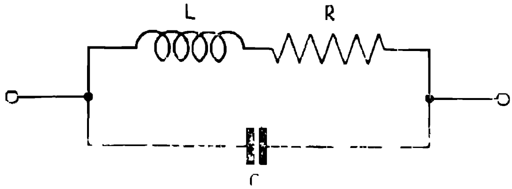


FIG. 181.—Parallel resonant circuit.

3. *Parallel circuit, with R assumed constant* (Q varies with ω).—

(a) Condition for resonance: —

$$\omega^2 = \frac{1}{LC} - \frac{R^2}{L^2} = \omega_0^2\left(1 - \frac{1}{Q^2}\right) \tag{62}$$

(b) Condition for maximum impedance :—

(i) If ω be varied :

$$\omega^2 = \frac{1}{LC}\sqrt{1 + \frac{2R^2C}{L}} - \frac{R^2}{L^2} = \omega_0^2\left[\sqrt{1 + \frac{2}{Q^2}} - \frac{1}{Q^2}\right] \tag{63}$$

(ii) If L be varied :

$$L = \frac{1}{2\omega^2 C} + \frac{1}{2\omega}\sqrt{\frac{1}{\omega^2 C^2} + 4R^2} = L_0 \cdot \tfrac{1}{2}\left[\sqrt{1 + \frac{4}{Q^2}} + 1\right] \tag{64}$$

(iii) If C be varied :

$$C = \frac{L}{R^2 + \omega^2 L^2} = C_0 \cdot \left(1 + \frac{1}{Q^2}\right)^{-1} \tag{65*}$$

(c) Condition for maximum current through L, with constant total current :—

(i) If ω be varied :

$$\omega^2 = \frac{1}{LC} - \frac{R^2}{2L^2} = \omega_0^2 \cdot \left(1 - \frac{1}{2Q^2}\right) \tag{66}$$

(ii) If L be varied :

$$L = \frac{1}{\omega^2 C} = L_0 \tag{67}$$

(iii) If C be varied :

$$C = \frac{L}{R^2 + \omega^2 L^2} = C_0 \cdot \left(1 + \frac{1}{Q^2}\right)^{-1} \tag{68*}$$

(d) Condition for maximum current through C, with constant total current :—

(i) If ω be varied :

$$\omega^2 = \omega_0^2 \cdot \tfrac{1}{2}\left[\sqrt{1 + \frac{2}{Q^2}} + 1\right] \tag{69}$$

(ii) If L be varied :

$$L = L_0 \cdot \tfrac{1}{2}\left[\sqrt{1 + \frac{4}{Q^2}} + 1\right] \tag{70}$$

---

* *i.e.*, as for resonance.

(iii) If C be varied :

$$C = \frac{1}{\omega^2 L} = C_0 \tag{71}$$

4. ***Parallel circuit, with Q assumed constant*** (R varies with ω).—

(*a*) Condition for resonance :—

$$\omega^2 = \frac{1}{LC} - \frac{R^2}{L^2} = \omega_0^2\left(1 - \frac{1}{Q^2}\right) \tag{72}$$

(*b*) Condition for maximum impedance :—

(i) If ω be varied :

$$\omega^2 = \frac{1}{LC}\left(1 + \frac{1}{Q^2}\right)^{-1} = \omega_0^2\left(1 + \frac{1}{Q^2}\right)^{-1} \tag{73}$$

(ii) If L be varied :

$$L = \frac{1}{\omega^2 C} = L_0 \tag{74}$$

(iii) If C be varied :

$$C = \frac{L}{R^2 + \omega^2 L} = C_0\left(1 + \frac{1}{Q^2}\right)^{-1} \tag{75}*$$

(*c*) Condition for maximum current through L, with constant total current :—

(i) If ω be varied :

$$\omega^2 = \frac{1}{LC}\left(1 + \frac{1}{Q^2}\right)^{-1} = \omega_0^2\left(1 + \frac{1}{Q^2}\right)^{-1} \tag{76}$$

(ii) If L be varied :

$$L = \frac{1}{\omega^2 C}\left(1 + \frac{1}{Q^2}\right)^{-1} = L_0\left(1 + \frac{1}{Q^2}\right)^{-1} \tag{77}*$$

(iii) If C be varied :

$$C = \frac{1}{\omega^2 L}\left(1 + \frac{1}{Q^2}\right)^{-1} = C_0\left(1 + \frac{1}{Q^2}\right)^{-1} \tag{78}*$$

(*d*) Condition for maximum current through C, with constant total current :—

(i) If ω be varied :

$$\omega^2 = \frac{1}{LC} = \omega_0^2 \tag{79}$$

(ii) If L be varied :

$$L = \frac{1}{\omega^2 C} = L_0 \tag{80}$$

(iii) If C be varied :

$$C = \frac{1}{\omega^2 L} = C_0 \tag{81}$$

---

* *i.e.*, as for resonance.

CHAPTER 5

AC CIRCUITS

GENERAL NETWORK THEOREMS

An " electrical network " may be defined as being *any electrical circuit containing impedances and generators.* A simple network may consist of a single closed circuit (or mesh), as in Fig. 182, whereas more complex networks may consist of a number of meshes that are interdependent, as in Fig. 183.

The current through, and the voltage across, any impedance of a network may be determined by the application of Ohm's and Kirchhoff's Laws, but, in the case of a complex network, the process

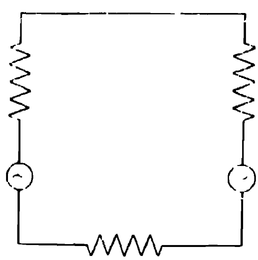


FIG. 182—Simple electrical network (single mesh).

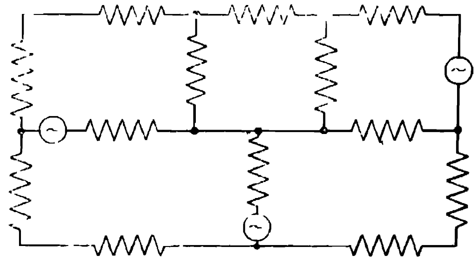


FIG. 183—More complex electrical network.

is lengthy and tedious, owing to the need for solving a large number of simultaneous equations. A number of network theorems have therefore been formulated to provide certain simplifications in the calculations.

Certain of these theorems have universal application, whereas others are restricted to circuits containing " linear " impedances. A " linear impedance " may be defined as *any impedance that obeys Ohm's Law ;* that is to say, any impedance for which the potential drop across it is proportional to the current flowing through it. Resistances, inductances and condensers fall into this category, whereas metal rectifiers and thermionic valves do not.

Superposition theorem

If a network of linear impedances contains more than one generator, the current flowing at any point is the vector sum of the currents that would flow at that point if each generator were considered separately

with all other generators replaced at the time by impedances equal to their internal impedances.

This follows from the linearity of Ohm's Law. For suppose the network consists of n meshes; Kirchhoff's Laws will give a set of n linear equations such as: –

$$\begin{aligned}
E_1 &= Z_{11}I_1 + Z_{12}I_2 + \ldots + Z_{1n}I_n \\
E_2 &= Z_{21}I_1 + Z_{22}I_2 + \ldots + Z_{2n}I_n \\
&\ldots\ldots \\
E_n &= Z_{n1}I_1 + Z_{n2}I_2 + \ldots + Z_{nn}I_n
\end{aligned}$$

These may be solved, giving a set of relations: —

$$\begin{aligned}
I_1 &= A_{11}E_1 + A_{12}E_2 + \ldots + A_{1n}E_n \\
I_2 &= A_{21}E_1 + A_{22}E_2 + \ldots + A_{2n}E_n \\
&\ldots\ldots \\
I_n &= A_{n1}E_1 + A_{n2}E_2 + \ldots + A_{nn}E_n
\end{aligned}$$

where the As are coefficients depending on the Zs, but independent of the Es and Is.

The theorem follows, from the linearity of these equations.

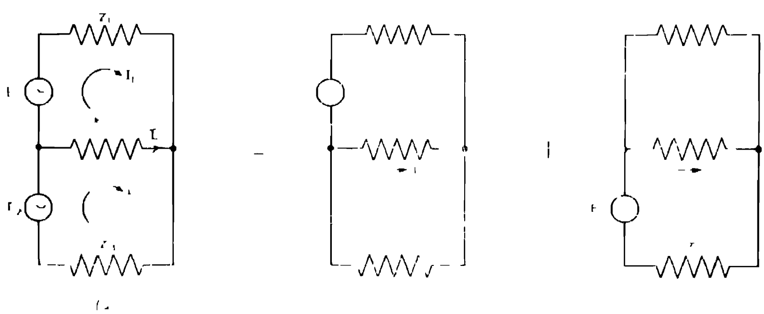


FIG. 184.- Equivalent circuits illustrating superposition theorem

Fig. 184 shows the superposition theorem applied to solve a simple problem. It is required to find the current I flowing in the impedance Z_2. This will first be solved by the application of Kirchhoff's Laws.

(*a*) *By Kirchhoff's Laws*.- Considering cyclic currents I_1 and I_2 as in Fig. 184*a*, the required current $I = I_2 - I_1$, and Kirchhoff's first law is satisfied.

By Kirchhoff's second law :—

$$I_1(Z_1 + Z_2) - I_2Z_2 = E_1$$

and

$$-I_1Z_2 + I_2(Z_2 + Z_3) = E_2$$

Solving these equations gives :—

$$I_1 = \frac{E_1(Z_2 + Z_3) + E_2Z_2}{Z_2Z_3 + Z_3Z_1 + Z_1Z_2}$$

$$I_2 = \frac{E_1Z_2 + E_2(Z_1 + Z_2)}{Z_2Z_3 + Z_3Z_1 + Z_1Z_2}$$

$$I_3 = I_2 - I_1 = \frac{E_1Z_2 + E_2(Z_1 + Z_2) - [E_1(Z_2 + Z_3) + E_2Z_2]}{Z_2Z_3 + Z_3Z_1 + Z_1Z_2}$$

i.e.
$$I_3 = \frac{E_2Z_1 + E_1Z_3}{Z_2Z_3 + Z_3Z_1 + Z_1Z_2}$$

(*b*) *By the superposition theorem* (see Fig. 184 *b* and *c*)

$$I_3 = i_1 + i_2$$

But
$$i_1 = \frac{E_1}{Z_1 + \dfrac{Z_2Z_3}{Z_2 + Z_3}} \times \frac{Z_3}{Z_2 + Z_3} = \frac{E_1Z_3}{Z_2Z_3 + Z_3Z_1 + Z_1Z_2}$$

and
$$i_2 = \frac{E_2}{Z_3 + \dfrac{Z_1Z_2}{Z_1 + Z_2}} \times \frac{Z_1}{Z_1 + Z_2} = \frac{E_2Z_1}{Z_2Z_3 + Z_3Z_1 + Z_1Z_2}$$

Therefore
$$I_3 = i_1 + i_2 = \frac{E_2Z_1 + E_1Z_3}{Z_2Z_3 + Z_3Z_1 + Z_1Z_2}$$

In this example there is little to choose between the two methods, but in a more complicated example the superposition method is often easier than the method using Kirchhoff's laws.

Thévenin's theorem

The current in a load impedance connected to two terminals a and b of a network of impedances and generators is the same as if this load impedance were connected to a simple constant-voltage generator, whose EMF is the open-circuit voltage measured across a and b and whose internal impedance is the impedance of the network looking back into the terminals a and b with all generators replaced by impedances equal to their internal impedances.

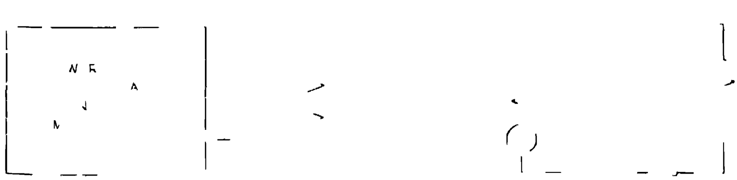


Fig. 185.—Equivalent circuits illustrating Thévenin's theorem

Fig. 185 shows Thévenin's theorem (sometimes also called "Pollard's theorem") in diagrammatic form. E_{oc} is the open-circuit voltage across terminals a b (i.e. with Z_l disconnected), and Z_{ab} is the impedance measured looking into terminals a b with all generators replaced by impedances equal to their internal impedances.

Example.—A network consisting of generators and impedances has two output terminals (*see* Fig. 186*a*). The following observations are made at these terminals:—

(*a*) Open circuit voltage 100 volts;

(*b*) current 2 amps with terminals short-circuited,

(*c*) current 1.77 amps in a 10 Ω resistance connected across the terminals.

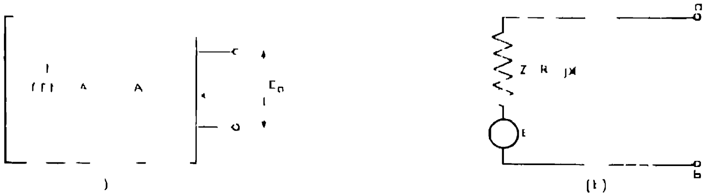


Fig. 186.—Example to illustrate use of Thevenin's theorem

Find the current that will flow in a 50 Ω resistance connected across the terminals.

The first step is to establish the equivalent circuit. This, by Thevenin's theorem, is the generator of Fig. 186b, where E_{oc} = 100 volts by (*a*) above.

To find $Z = R + jX$ use the data (*b*) and (*c*).

(*b*) gives
$$\frac{100}{\sqrt{R^2 + X^2}} = 2$$

$$\therefore \quad \frac{100^2}{4} = R^2 + X^2 \qquad (i)$$

(*c*) gives
$$\frac{100}{\sqrt{(R + 10)^2 + X^2}} = 1.77$$

$$\therefore \quad \frac{100^2}{3.132} = R^2 + X^2 + 20R + 100 \qquad (ii)$$

From (i) and (ii)
$$20R + 100 = \frac{100^2}{3.132} - \frac{100^2}{4}$$
$$= 100^2 \, \frac{0.87}{4 \times 3.13}$$
$$R = \frac{500 \times 0.87}{4 \times 3.13} - 5$$
$$= 35 - 5$$
$$= 30 \text{ ohms}$$

From (i) $\quad R^2 + X^2 = 2500$

$\therefore \quad X^2 = 2500 - 900 = 1600$

$\therefore \quad X = 40$ ohms

$$\therefore \quad Z = 30 + j40$$

With a resistive load of 50 Ω,

$$I = \frac{100}{\sqrt{80^2 + 40^2}} = \frac{100}{40\sqrt{5}} = 1 \cdot 12 \text{ amps} \quad \textit{Ans}$$

Norton's theorem

The current in a load impedance connected to two terminals a and b of a network consisting of generators and impedances is the same as if this load impedance were connected to a constant-current generator, whose generated current is equal to the short circuit current measured at a b, and having infinite internal impedance but placed in parallel with an impedance equal to the impedance of the network looking back into the terminals a and b with all generators replaced by impedances equal to their internal impedances

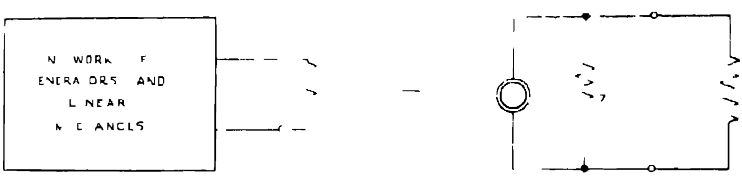


FIG 187 —Equivalent circuits illustrating Norton's theorem

This theorem is similar to Thévenin's theorem in that it enables a complicated network to be replaced by a single generator and impedance. In this case however the generator is of the constant-current type and the impedance is in shunt with it whilst in the case of Thévenin's theorem the equivalent generator is of the constant-voltage type in series with an impedance The equivalent circuits given by Thévenin's and Norton's theorem yield exactly the same current and voltage in the load impedance, and are therefore effectively identical to one another In any particular problem either theorem can therefore be used

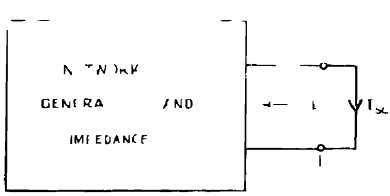


FIG 188 —Derivation of constant current generator

In most cases Thévenin's theorem is the easier to apply, although when the network impedance is high compared with the load impedance, the constant-current generator concept (Norton's theorem) may simplify calculations—as in the equivalent circuit for a pentode valve amplifier, *see* page 361.

Fig. 187 shows Norton's theorem in diagrammatic form. I_{sc} is the current that flows when terminals a and b are short-circuited and Z_{ab} is the impedance measured looking into terminals a b with all generators replaced by impedances equal to their internal impedances (see Fig. 188).

Compensation theorem

Any impedance in a network can be replaced by a generator of zero internal impedance and of EMF equal to the instantaneous potential difference across the replaced impedance.

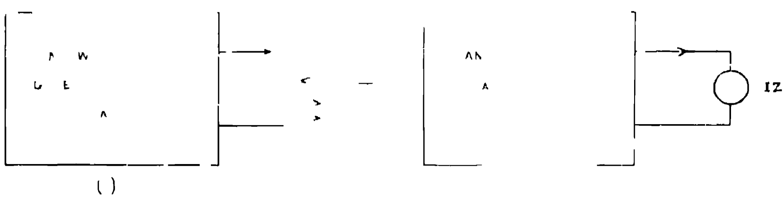


Fig. 189. Equivalent networks illustrating compensation theorem

Fig. 189 shows a network of impedances and generators together with the particular impedance Z that is to be replaced considered as the load. The network may be completely solved by using Kirchhoff's Laws. In particular the equation for the right hand mesh will be

$$\sum Z_1 I_1 + ZI = \sum E_1$$

where the summation extends over the number of unspecified impedances in the right hand mesh.

If the circuit of Fig. 189*a* is solved by Kirchhoff's Laws the equations will be exactly the same as for Fig. 189*b* with the exception of the equation for the right hand mesh which becomes —

$$\sum Z_1 I_1 = -ZI + \sum E_1$$

Thus all the equations are identical for the two networks and so also are the currents and voltages throughout the two networks; that is the networks are equivalent.

It will be noticed that there is no restriction in this theorem on the types of impedance in the network; the impedances may in fact be linear or non linear.

Maximum power transfer theorem

This theorem states that, *given a generator with internal impedance $Z_g \angle \phi$ the maximum power will be obtained from it if a load having the conjugate impedance $Z_g \angle -\phi$ is connected across it. If the modulus alone can be varied the power will be a maximum if the moduli of the load and generator impedances are equal, irrespective of the value of ϕ.* This will now be proved.

Let a load of impedance $Z_L \angle \theta$ be connected across the generator as in Fig. 190.

The problem is to determine the values of Z_L and θ to give maximum power in the load.

The first step is to calculate the power in the load. It has been seen that the vector $Z_L \angle \theta$ is equivalent to the vector $Z_L(\cos\theta + j\sin\theta)$ (see Chapter 2). Thus the total impedance is

$$Z = Z_G \angle \varphi + Z_L \angle \theta = (Z_G\cos\varphi + Z_L\cos\theta) + j(Z_G\sin\varphi + Z_L\sin\theta) \qquad (1)$$

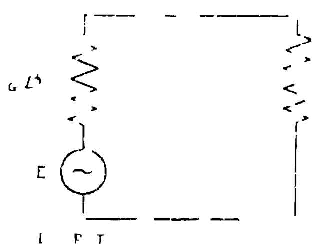


FIG. 190.—Maximum power transfer from a generator connected to a load

$$\therefore |Z|^2 = (Z_G\cos\varphi + Z_L\cos\theta)^2 + (Z_G\sin\varphi + Z_L\sin\theta)^2$$
$$= Z_G^2(\cos^2\varphi + \sin^2\varphi) + Z_L^2(\cos^2\theta + \sin^2\theta) + 2Z_GZ_L(\cos\varphi\cos\theta + \sin\varphi\sin\theta)$$
$$= Z_G^2 + Z_L^2 + 2Z_GZ_L\cos(\varphi - \theta) \qquad (2)$$

$$|I| = \frac{|E|}{|Z|}$$

and—

$$P = |I|^2 \text{ (the resistive part of the load impedance)}$$
$$= |I|^2 Z_L \cos\theta$$
$$\therefore P = \frac{|E|^2}{|Z|^2} Z_L\cos\theta$$
$$= \frac{|E|^2 Z_L\cos\theta}{|Z_G^2 + Z_L^2 + 2Z_GZ_L\cos(\varphi-\theta)|} \qquad (3)$$

Consider first the variation of Z_L with θ constant. P will be a maximum when—

$$\frac{dP}{dZ_L} = 0$$

$$\text{Now } P = \frac{|E|^2 Z_L\cos\theta}{Z_G^2 + Z_L^2 + 2Z_GZ_L\cos(\varphi-\theta)} = \frac{|E|^2\cos\theta}{\frac{Z_G^2}{Z_L} + Z_L + 2Z_G\cos(\varphi-\theta)}$$

$$\therefore \quad \frac{dP}{dZ_L} = \frac{-|E|^2 \cos\theta \left[\dfrac{-Z_G^2}{Z_L^2} + 1 \right]}{\left[\dfrac{Z_G^2}{Z_L} + Z_L + 2Z_G \cos(\varphi - \theta) \right]^2} \tag{4}$$

This is equal to zero if $Z_L^2 = Z_G^2$ and since Z_L is positive, this gives $Z_L = Z_G$.

As $\dfrac{dP}{dZ_L}$ is positive when $Z_L < Z_G$ and negative when $Z_L > Z_G$, this gives a maximum value to P. Hence the power is a maximum when the moduli of the generator and load impedances are equal.

Now consider variations of θ with Z_L constant.

$$P = \frac{|E|^2 Z_L \cos\theta}{Z_G^2 + Z_L^2 + 2Z_G Z_L \cos(\varphi - \theta)}$$

$$\frac{dP}{d\theta} = \frac{|E|^2 Z_L [-\sin\theta\{Z_G^2 + Z_L^2 + 2Z_G Z_L \cos(\varphi - \theta)\} - \cos\theta \, 2Z_G Z_L \sin(\varphi - \theta)]}{[Z_G^2 + Z_L^2 + 2Z_G Z_L \cos(\varphi - \theta)]^2}$$

$$= \frac{-|E|^2 Z_L [(Z_G^2 + Z_L^2)\sin\theta + 2Z_G Z_L \sin\varphi]}{[Z_G^2 + Z_L^2 + 2Z_G Z_L \cos(\varphi - \theta)]^2} \tag{5}$$

This is zero if $\sin\theta = -\sin\varphi \dfrac{2Z_G Z_L}{Z_G^2 + Z_L^2}$ and, as before, this gives a maximum value to P.

The power will be an absolute maximum if both conditions are satisfied simultaneously, i.e. $Z_L = Z_G$ and $\sin\theta = -\sin\varphi$, or $\theta = -\varphi$, thus the load for maximum power is $Z_G \angle -\varphi$.

Writing the generator impedance as in the form $R + jX$, the load for maximum power is $R - jX$, that is, the generator and load impedances are conjugates.

It should be noted that this theorem applies to variation in *load* impedance. If the generator impedance is variable, maximum power will be transferred when $Z_G = 0 - jX_L$.

THE DECIBEL

Power ratios

Line communication is concerned with the transmission of AC power from one point to another and the various lines and pieces of equipment that constitute a communication system introduce gains and losses of power

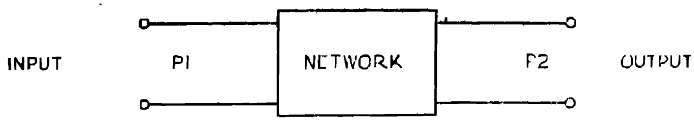


FIG. 191.—Network connecting a generator and a load

Consider a network connecting a generator to a load. Let the input power be P_1 and the output power be P_2. The ratio of output power to input power is then $\frac{P_2}{P_1}$. The network may introduce a loss, in which case $\frac{P_2}{P_1}$ will be less than unity, or it may introduce a gain (*e.g.* an amplifier), in which case $\frac{P_2}{P_1}$ will be greater than unity.

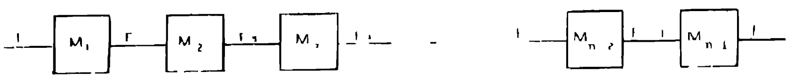


FIG. 192.—Series of networks in tandem

If a number of such networks are connected in tandem as in Fig. 192, and the individual power ratios are known, the overall power ratio $\frac{P_n}{P_1}$ is obtained by multiplying together the individual power ratios. This follows from the fact that:—

$$\frac{P_n}{P_1} = \frac{P_2}{P_1} \times \frac{P_3}{P_2} \times \frac{P_4}{P_3} \times \dots \times \frac{P_n}{P_{n-1}}$$
$$= M_1 \times M_2 \times M_3 \times \dots \times M_n,$$

where M_1, *etc.*, are the individual power ratios.

Example.—

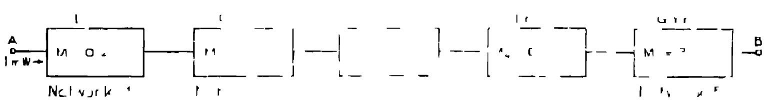


FIG. 193

Consider Fig. 193. Five networks are inserted in tandem between A and B. The individual power ratios are:—

$$M_1 = 0\cdot215 \quad \text{(a loss)}$$
$$M_2 = 20\cdot3 \quad \text{(a gain)}$$
$$M_3 = 0\cdot0246 \quad \text{(a loss)}$$
$$M_4 = 0\cdot251 \quad \text{(a loss)}$$
$$M_5 = 25\cdot2 \quad \text{(a gain)}$$

Find the power at B if 1 mW of power is applied to A.

$$\text{Overall power ratio} = 0{\cdot}215 \times 20{\cdot}3 \times 0{\cdot}0246 \times 0{\cdot}251 \times 25{\cdot}2$$
$$= 0{\cdot}679 \text{ (a loss)}$$

Thus if 1 mW is applied to A the output power at B will be 0·679 mW. *Ans.*

Logarithmic units

In a complex system containing a large number of component circuits each contributing a gain or loss, calculation of the overall power ratio may become extremely laborious. To simplify this calculation the individual power ratios are expressed in a logarithmic unit enabling addition to be employed in place of multiplication. The logarithmic unit employed is the "decibel" (abbreviated to db) and power gain or loss D of a network expressed in this unit is defined as

$$D = 10 \log_{10} \frac{P_2}{P_1} \tag{6}$$

where P_2 = output power
and P_1 = input power

If $\frac{P_2}{P_1}$ is less than unity, then $10 \log_{10} \frac{P_2}{P_1}$ will be negative. A negative sign thus indicates a power loss and a positive sign a gain.

It should be noted that since

$$10 \log_{10} \frac{P_2}{P_1} = -10 \log_{10} \frac{P_1}{P_2} \tag{7}$$

the numerical answer will be the same no matter whether $\frac{P_2}{P_1}$ or $\frac{P_1}{P_2}$ is considered, but to obtain the correct sign $\frac{P_2}{P_1}$ must be considered.

The following examples given in Table IX show how power ratios can be expressed in decibels.

TABLE IX

Power ratios expressed in decibels

Input	Output	Ratio	Gain in db (negative sign indicates a loss)
2 mW	2000 mW	1000	$10 \log_{10} 1000 = 10 \times 3 = 30$ db
3 mW	600 mW	200	$10 \log_{10} 200 = 10 \times 2{\cdot}301 = 23{\cdot}01$ db
5 mW	500 mW	100	$10 \log_{10} 100 = 10 \times 2 = 20$ db
20 mW	2000 mW	100	$10 \log_{10} 100 = 10 \times 2 = 20$ db
2 mW	20 mW	10	$10 \log_{10} 10 = 10 \times 1 = 10$ db
40 mW	200 mW	5	$10 \log_{10} 5 = 10 \times 0{\cdot}699 = 6{\cdot}99$ db
3 W	6 W	2	$10 \log_{10} 2 = 10 \times 0{\cdot}301 = 3{\cdot}01$ db
10 mW	12·6 mW	1·26	$10 \log_{10} 1{\cdot}26 = 10 \times 0{\cdot}10 = 1$ db
500 mW	5 mW	0·01	$10 \log_{10} (0{\cdot}01) = 10 \times \bar{1} \ldots 100$; $10 \times (-2) = -20$ db
100 mW	21·6 mW	0·216	$10 \log_{10} 0{\cdot}216 = 10(\bar{1}{\cdot}335) = 10(-1+0{\cdot}335) = -6{\cdot}65$ db

Expression of absolute power level using the decibel notation

The decibel is fundamentally a unit of power *ratio* and not of absolute power, but if some standard reference level of power be assumed, then any absolute power can be expressed as so many db above or below this reference standard. While various other standards may occasionally be encountered, the standard adopted in Britain and America is 1 mW (0·001 W). Using this

TABLE X

Absolute power expressed in dbm

Powers expressed in decibel referred to 1 mW

1	$\mu\mu$W	− 90 dbm	1 mW	0 dbm
10	$\mu\mu$W	80 ,,	2 mW	+ 3 ,,
100	$\mu\mu$W	− 70 ,,	4 mW	+ 6 ,,
0·001	μW	− 60 ,,	5 mW	+ 7 ,,
0·01	μW	− 50 ,,	8 mW	+ 9 ,,
0·1	μW	− 40 ,,	10 mW	+ 10 ,,
1·0	μW	30 ,,	20 mW	+ 13 ,,
2	μW	− 27 ,,	40 mW	+ 16 ,,
4	μW	− 24 ,,	50 mW	+ 17 ,,
5	μW	23 ,,	80 mW	+ 19 ,,
8	μW	− 21 ,,	100 mW	+ 20 ,,
10	μW	20 ,,	200 mW	+ 23 ,,
20	μW	17 ,,	400 mW	+ 26 ,,
40	μW	14 ,,	500 mW	+ 27 ,,
50	μW	− 13 ,,	800 mW	+ 29 ,,
80	μW	11 ,,	1000 mW } 1 W }	+ 30 ,,
100	μW	10	2 W	+ 33 ,,
200	μW	7 ,,	4 W	+ 36 ,,
400	μW	4	5 W	+ 37 ,,
500	μW	− 3 ,,	8 W	+ 39 ,,
800	μW	− 1	10 W	+ 40 ,,
			100 W	+ 50 ,,
			1 kW	+ 60 ,,
			10 kW	+ 70 ,,
			100 kW	+ 80 ,,

standard any power P can be expressed as $10 \log_{10}\left(\frac{P}{1\ \mathrm{mW}}\right)$ db referred to 1 mW. Thus one can express 1 watt as $10 \log_{10}\left(\frac{1000}{1}\right) = +30$db above the standard 1 mW or "+30 db with respect to 1 mW". Similarly, 5 μW can be expressed as $10 \log_{10}\left(\frac{5}{1000}\right) = -23$ db with respect to 1 mW (*i.e.*, 23 db below 1 mW). The expression "db with respect to 1 mW" or "db referred to 1 mW" is usually abbreviated to "dbm"; other abbreviations sometimes met are "db wrt 1 mW", "db ref 1 mW", and "vu" (voice unit).

Thus $\quad 20 \text{ dbm} \equiv 20 \text{ db wrt } 1 \text{ mW} \equiv 20 \text{ db ref } 1 \text{ mW} \equiv 20 \text{ vu}$

$$= 20 \text{ db with respect to } 1 \text{ mW}$$
$$= 100 \text{ mW}.$$

Conversion of decibels to power ratios

The conversion of power ratios expressed in decibels (and of powers expressed in dbm) to actual power ratios (and actual powers) is effected by exactly the reverse process from that used for expressing power ratios in decibels (and actual powers in dbm)

$$10 \log_{10} \frac{P_2}{P_1} = D \qquad \text{(8) (quoted)}$$

$$\therefore \qquad \log_{10} \frac{P_2}{P_1} = \frac{D}{10}$$

$$\therefore \qquad \frac{P_2}{P_1} = \text{antilog}\left(\frac{D}{10}\right) \qquad (9)$$

Example 1 —

What power in watts is represented by 25 dbm ?

$$\frac{P}{1 \text{ mW}} = \text{antilog}\, \frac{25}{10} \qquad \text{antilog } 2.5 = 316.2$$

$$P = 316{\cdot}2 \text{ mW}$$
$$= 0.316 \text{ W} \quad \textit{Ans}$$

Example 2.—

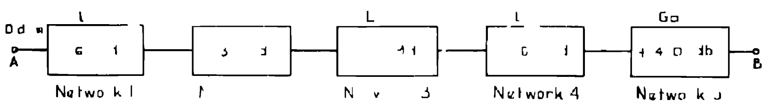


Fig. 194

Consider Fig. 194 Five networks are inserted in tandem between A and B The decibel gains and losses of the individual networks are —

Network 1	− 6 68 db
Network 2	+ 13 08 db
Network 3	— 16 09 db
Network 4	— 6 00 db
Network 5	+ 14 01 db

Find the power at B if 0 dbm is applied to A

Total decibel gain or loss $\quad$ 6 68+13 08—16 09—6 00+14 01

$= -1\ 68$ db

Power at B $\quad 0 - 1\ 68$ dbm

$-1\ 68$ dbm $\quad$ *Ans.*

This is the decibel approach to the example given on page 239 , it should be noted that — 1 68 dbm corresponds to a power of 0·679 milliwatts.

Current and voltage ratios

When it is desired to compare the powers developed in two *equal resistors* it is sufficient to measure the two voltages or the two currents; then the power ratio in decibels is equal to twenty times the logarithm (to base 10) of the current or voltage ratio.

Consider two equal resistors of R ohms, carrying currents of RMS values I_1 and I_2, and having voltages across them of RMS values E_1 and E_2 respectively. Then the powers developed in these two resistors are :–

$$P_1 = E_1 I_1 = R \cdot I_1^2 = \frac{1}{R} \cdot E_1^2$$

and

$$P_2 = E_2 I_2 = R \cdot I_2^2 = \frac{1}{R} \cdot E_2^2$$

The ratio between these two powers is therefore :—

$$\frac{P_2}{P_1} = \frac{E_2 I_2}{E_1 I_1} = \left(\frac{I_2}{I_1}\right)^2 = \left(\frac{E_2}{E_1}\right)^2$$

Or, expressing this in decibels,

$$D = 10 \log_{10}\left(\frac{P_2}{P_1}\right) = 10 \log_{10}\left(\frac{I_2}{I_1}\right)^2$$

$$= 20 \log_{10}\left(\frac{I_2}{I_1}\right) \qquad (10)$$

and

$$D = 10 \log_{10}\left(\frac{P_2}{P_1}\right) = 10 \log_{10}\left(\frac{E_2}{E_1}\right)^2$$

$$= 20 \log_{10}\left(\frac{E_2}{E_1}\right) \qquad (11)$$

Thus the power ratio in decibels is equal to 20 $\log_{10}$ (current ratio) = 20 $\log_{10}$ (voltage ratio), *provided that the two resistances are equal* through which the two currents I_1 and I_2 (or across which the two voltages E_1 and E_2) are measured.

Currents through impedances.—When it is required to compare the powers in two impedances by measurement of the currents through them, it is desirable that their resistive components be equal since, in this case, the power ratio will be equal to the (current ratio)^2. This may be verified as follows :—

Let the two impedances be : –

$$Z_1 = |Z_1| \underline{/\varphi_1} = R_1 + jX_1$$

and

$$Z_2 = |Z_2| \underline{/\varphi_2} = R_2 + jX_2$$

Let the magnitudes of the currents flowing in Z_1 and Z_2 be I_1 and I_2 respectively.

The power ratio is :—

$$\frac{P_2}{P_1} = \frac{I_2^2 R_2}{I_1^2 R_1} \qquad (12)$$

Hence when $\qquad R_1 = R_2$

$$\frac{P_2}{P_1} = \left(\frac{I_2}{I_1}\right)^2.$$

Voltages across impedances — When considering the voltage across two impedances the (voltage ratio)^2 will be equal to the power ratio if the conductive components be equal. This may be verified as follows:—

Let the impedances be:—

$$Z_1 = |Z_1| \angle \varphi_1 = R_1 + jX_1$$

and $\qquad Z_2 = |Z_2| \angle \varphi_2 = R_2 + jX_2$

Let their admittances be

$$Y_1 = |Y_1| \angle -\varphi_1 = G_1 + jB_1$$

and $\qquad Y_2 = |Y_2| \angle -\varphi_2 = G_2 + jB_2$

Power $\qquad P_1 = I_1^2 R_1 = \frac{E_1^2}{Z_1^2} Z_1 \cos \varphi_1$

$$= \frac{E_1^2}{Z_1} \cos \varphi_1$$

$$= E_1^2 Y_1 \cos \varphi_1$$

$$= E_1^2 G_1$$

Similarly $\qquad P_2 = E_2^2 G_2$

The power ratio is

$$\frac{P_2}{P_1} = \frac{E_2^2 G_2}{E_1^2 G_1} \tag{13}$$

Hence when $\qquad G_1 = G_2$

$$\frac{P_2}{P_1} = \left(\frac{E_2}{E_1}\right)^2$$

TRANSFORMERS

Two or more coils possessing mutual inductance form a transformer. At audio and power frequencies iron cores are generally used and the mutual inductance is large. At higher frequencies iron cores cannot be used and the coupling is smaller; such transformers are most conveniently considered as coupled circuits. The behaviour of iron cored transformers is most easily seen by considering first a "perfect" or 'ideal' transformer, that is, one with no losses. The losses that occur in practical transformers can then be considered separately.

The "perfect" transformer

Consider first a transformer with zero winding resistance, an infinite primary inductance and such that all flux produced by the primary cuts the secondary and vice versa. This is known as a perfect transformer. Its behaviour may be explained by a vector diagram. As the flux is the only link between primary

and secondary, it is chosen as the reference vector (it can be represented by a vector as it varies sinusoidally).

Conditions on no-load

First consider the transformer with a constant alternating voltage E_1 applied to the primary and the secondary open circuited—*i.e.* off load (*see* Fig. 195*a*). Let the primary and secondary turns be T_1 and T_2 respectively. The current that flows will be infinitely small due to the large inductance of the primary; this will be 90° behind the voltage E_1. The magnetic flux Φ in the iron core will be in phase with this current, hence Φ is 90° behind E_1. The flux, however, will still be finite and the vector diagram is as shown in Fig. 195*b*.

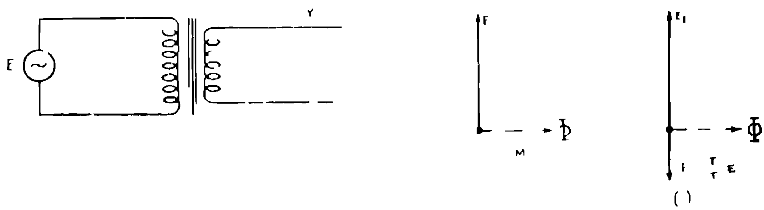


FIG. 195.—Transformer on no-load (open circuit)

The next result is one of fundamental importance upon which is based the whole of transformer theory. It is as follows:—

If the applied voltage E_1 is constant, then the flux Φ is constant.

Since sine curves are being dealt with, the term "constant" refers to the amplitude of the waveform, *i.e.* the moduli of the corresponding vectors.

As it is always assumed that the applied voltage is constant, it will be seen that the flux must remain constant whatever load may be connected to the secondary. The proof of the statement just made is as follows:

If there is no resistance in the transformer, the applied EMF must be exactly equal and opposite to the back EMF developed across the primary. Hence, as E_1 is constant, the back EMF must be constant. But the back EMF is proportional to the rate of change of flux; the flux must therefore be such that its rate of change is a sine wave of constant amplitude. Thus the flux itself must be a sinusoidal waveform of constant amplitude, *i.e.* the flux is constant.

Returning to the perfect transformer with voltage E_1 applied to the primary, the secondary voltage E_2 can be calculated. For E_2 is produced by the flux cutting the secondary turns, and as the flux is the same for primary and secondary, the primary and secondary induced voltages will be proportional to the primary

turns T_1 and the secondary turns T_2 respectively. But the primary induced voltage is the back-EMF in the primary, which is equal and opposite to E_1. Hence the secondary voltage is of magnitude E_2 such that

$$\frac{E_1}{E_2} = \frac{T_1}{T_2} \tag{14}$$

Also, E_2 is in phase with the back-EMF in the primary, *i.e.* 180° out of phase with the applied primary voltage. This result is also true when any load is connected across the secondary. The vector diagram is shown in Fig. 195*c*.

The derivation of the name "transformer" will now be clear: by choosing a suitable ratio for $\frac{T_1}{T_2}$ we can transform an alternating voltage to any other required voltage of the same frequency.

Conditions on load

It has been shown so far that as soon as a voltage is applied to the primary a voltage appears across the secondary, and a small amount of primary current flows (infinitely small if the transformer is perfect). It now remains to study the action when a load impedance is connected across the secondary. Assume that this

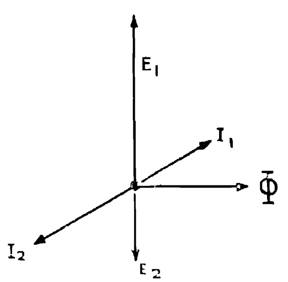


FIG. 196.—Vector diagram for perfect transformer on load.

impedance is given; the secondary voltage is constant and known, and hence the magnitude and phase of the secondary current can be calculated.

Insert this on the vector diagram; Fig. 196 shows the case for a load with a positive reactance, so that I_2 lags behind E_2. It is here that the constancy of the flux Φ must be taken into account, for the effect of this current I_2 on the flux will be considerable. Some action must take place to nullify this effect: in fact, a primary current I_1 flows.

If I_1 is to have the opposite effect to I_2, it must be 180° out of phase. Its magnitude is determined by the fact that the two effects are to be equal. Now the effect of a current I on flux is proportional to $I \times T$, where T is the number of turns through which it flows. As the effect of I_1 and I_2 must be equal, I_1T_1 and I_2T_2 must be equal.

$$I_1T_1 = I_2T_2$$

$$\frac{I_1}{I_2} = \frac{T_2}{T_1} \tag{15}$$

In other words, the current ratio is the inverse of the turns ratio, *i e.* the winding with the fewer turns carries the larger current. Fig. 196 shows the vector diagram

Note that the angle between E_2 and I_2 is equal to the angle between E_1 and I_1, *i e* if the load on the secondary takes a lagging current, the transformer will take a lagging current from the supply This can be put in another way by saying that the power factor is the same for primary and secondary

It has been shown that $\frac{E_1}{E_2} = \frac{T_1}{T_2}$ and that $\frac{I_2}{I_1} = \frac{T_1}{T_2}$, hence $\frac{E_1}{E_2} = \frac{I_2}{I_1}$

$$\therefore \qquad E_1 I_1 = E_2 I_2$$

The input power $= E_1I_1 \cos \varphi_1$ where $\cos \varphi_1$ is the input power factor, and the output power $= E_2I_2 \cos \varphi_2$ where $\cos \varphi_2$ is the output power factor But $\varphi_1 = \varphi_2$ so that $\cos \varphi_1 = \cos \varphi_2$ and it has just been proved that $E_1I_1 = E_2I_2$

$$\therefore \qquad \textit{Input power} = \textit{Output power} \tag{16}$$

Thus the perfect transformer introduces no loss, and the efficiency is 100 per cent In practice efficiencies of 95 per cent. can easily be obtained in power transformers and audio frequency transformers frequently have a loss of less than $\frac{1}{2}$ db

Impedance transformation

With the secondary off load the primary of a perfect transformer takes no current, that is to say its input impedance is infinite When a load is connected to the secondary, the input or

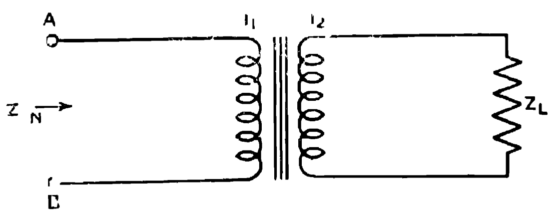


FIG 197 —Impedance transforming property of a transformer.

primary impedance Z_{IN} is not infinite, but depends upon the load Z_L in the secondary (*see* Fig 197) The relationship between Z_{IN} and Z_L is most important, and will now be derived.

From equations 14 and 15,

$$\frac{E_1}{E_2} = \frac{T_1}{T_2} \text{ and } \frac{I_2}{I_1} = \frac{T_1}{T_2}$$

$$\frac{L_1 I_2}{L_2 I_1} = \frac{T_1^2}{T_2^2}$$

$$\frac{E_1/I_1}{E_2/I_2} = \frac{T_1^2}{T_2^2}$$

But $\frac{E_1}{I_1} = Z_{in}$ and $\frac{E_2}{I_2} = Z_L$

$$\frac{Z_{in}}{Z_L} = \frac{T_1^2}{T_2^2}$$

$$Z_{in} = \frac{T_1^2}{T_2^2} Z_L \qquad (17)$$

This means that the impedance Z_{in} across AB in Fig. 197 will be $\frac{T_1^2}{T_2^2} Z_L$. It should be noted that the transformer *alters the modulus of an impedance but not its angle*. Thus if Z_L is a condenser the impedance across AB will be capacitative.

If T_1 is greater than T_2 the input impedance Z_{in} will be greater than Z_L; if T_1 is less than T_2, the input impedance Z_{in} will be less than Z_L. Note that the impedance ratio is the *square* of the turns ratio.

This property of a transformer is known as "impedance transformation". It is very useful for connecting together two circuits (e.g. lines) of different impedance to satisfy the condition for maximum power transference.

From this result it can be seen that a series impedance Z in the secondary can be transferred to the primary as an impedance $Z \frac{T_1^2}{T_2^2}$ without affecting the behaviour of the circuit, e.g. the two circuits in Fig. 198 are equivalent as far as A.C. is concerned.

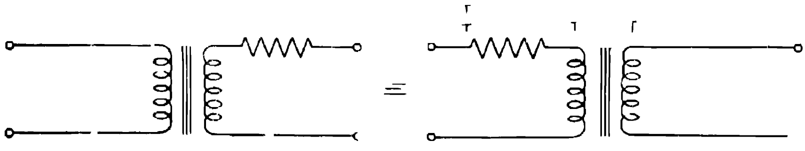
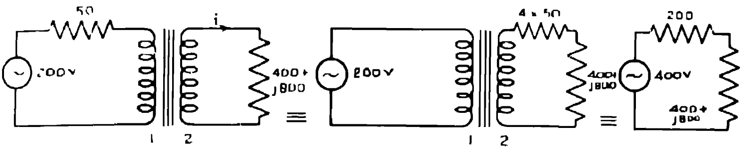


Fig. 198.—Secondary impedance transferred to primary.

In the same way an impedance could if desired be transferred from primary to secondary by multiplying by $\frac{T_2^2}{T_1^2}$. This method often simplifies the solution of problems involving transformers.

Example 1.—Find the power dissipated in the load of the transformer shown in Fig. 199*a*.

This will be achieved by finding the secondary current i and using the formula $P = i^2R = 400i^2$. To find i, the equivalent circuit of Fig. 199*b* is used.



(*a*) (*b*) (*c*)

FIG. 199.

The 50 ohms has been transferred to the secondary as $4 \times 50 = 200$ ohms. The transformer with 200v on the primary is equivalent to a generator of 400v, by Thévenin's theorem, so that a further equivalent circuit is as shown in Fig. 199*c*.

Total impedance $Z = 600 + j800 = 200\,(3 + j4)$

$$|Z| = 200\sqrt{3^2 + 4^2} = 1000 \text{ ohms}$$

$$\therefore \qquad i = \frac{400}{1000} = 0{\cdot}4 \text{ amps}$$

$$\therefore \qquad P = 400 \times i^2 = 0{\cdot}16 \times 400 = 64 \text{ watts.}$$

Example 2.—Find the equivalent input capacity of the transformer in Fig. 200.

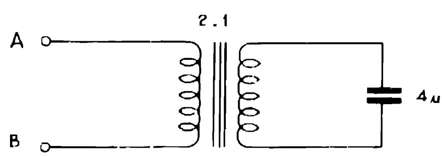


FIG. 200.

It is easy to be caught out by a simple example such as this, by saying that the answer is 16μF. This, of course, is *not* the correct answer.

The impedance across AB is $2^2 = 4$ times the impedance of the 4μF condenser. To *multiply* the *impedance* of a condenser by 4, its *capacity* must be *divided* by four. The correct answer is therefore 1μF.

Transformers and maximum power transfer

The maximum power transfer theorem stated (p. 236) that, to obtain maximum power from a generator of internal impedance $Z_g \angle \varphi$, a load of impedance $Z_g \angle -\varphi$ must be connected; and that if the two angles cannot be made equal and opposite, the maximum power under these circumstances will be obtained when the moduli of the two impedances are equal.

This does not answer the problem in its practical form: usually

the impedance of the generator and of the load are both given, and it is required to connect the two together to obtain maximum power transfer. This process is known as "impedance matching", and is carried out by using a transformer of suitable turns ratio. It was shown above that a transformer alters the modulus of an impedance by a factor $\frac{T_1^2}{T_2^2}$ but does not affect the angle. Hence if the impedances of the load and of the generator are given, a transformer can be used to satisfy the first condition of the maximum power transfer theorem (moduli equal), but not the second (angles equal and opposite). If, as is often the case, the two impedances are resistive, the second condition is automatically satisfied, as both angles are equal to zero.

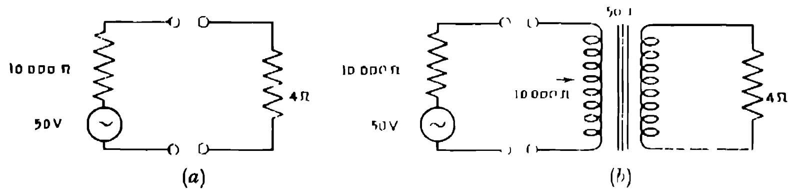


FIG. 201.—Example to illustrate impedance matching, using a transformer.

Fig. 201 illustrates a typical example: the load and generator are given in Fig. 201*a*.

If the connection was made direct, the power developed in the 4 ohms would be negligible. A transformer is therefore inserted; if the impedances are to match, $\frac{T_1^2}{T_2^2}$ will have to equal $\frac{10,000}{4}$;

$$\therefore \quad \frac{T_1^2}{T_2^2} = \frac{10,000}{4} = 50^2$$

This transforms the 4 ohms up to 10,000, as shown in Fig. 201*b*, and maximum power will be transferred.

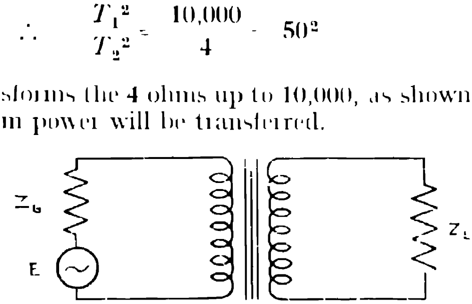


FIG. 202. Impedance matching, using a transformer (general case).

If the generator impedance is Z_G, and the load impedance is Z_L (*see* Fig. 202), the turns ratio will be given by:—

$$\frac{T_1}{T_2} = \sqrt{\frac{|Z_G|}{|Z_L|}} \tag{18}$$

Example.—

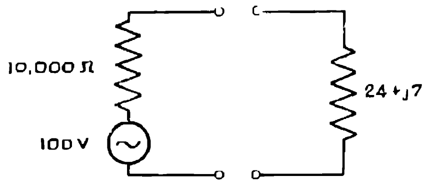


FIG. 203.

The generator and load are as given in Fig. 203. Calculate :—

(*a*) The power in the load if connected directly to the generator.

(*b*) The turns ratio of transformer for maximum power transfer, and the power so transferred.

(*c*) The power that would be transferred if the second condition of the maximum power transfer theorem could be satisfied (*i.e.* angles equal and opposite).

(*a*) Total impedance = 10,024 + j7. To slide-rule accuracy, this is equal to 10,000.

$$\therefore \qquad I = \frac{100}{10^4}$$

$$= \frac{1}{100} \text{ amp}$$

$$\therefore \qquad P = I^2R = \frac{1}{10^4} \times 24 \text{ watts}$$

$$= 2{\cdot}4 \text{ milliwatts}, \quad Ans.$$

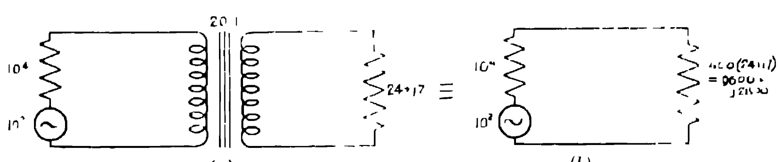


FIG. 204.

(*b*) The modulus of the load impedance Z_L is :—

$$|Z_L| = \sqrt{24^2 + 7^2}$$

$$= 25$$

$$\therefore \qquad \frac{T_1{}^2}{T_2{}^2} = \frac{10^4}{25}$$

$$\therefore \qquad \frac{T_1}{T_2} = 20 : 1$$

To find the power transferred, find the equivalent circuit ; this is shown in Fig. 204.

Note that the transferred impedance is *not* 10^4 ohms—although its modulus will be 10^4 ohms. It must be transferred as a vector, *i.e.* as $9600 + j \,.\, 2800$.

$$\text{Total impedance} = 10^4 + 9600 + j \,.\, 2800 = 19{,}600 + j \,.\, 2800$$
$$= 2800\,(7 + j.1)$$

$$\therefore \qquad I = \frac{100}{2800 \,.\, |7 + j.1|} = \frac{1}{28\sqrt{50}} \text{ amp}$$

$\therefore$ Power transferred

$$P = I^2R = \frac{1}{28^2 \times 50} \times 9600 = \frac{12}{49} \text{ watts}$$

$$\textit{i.e.} \qquad P = 245 \text{ mW} \quad \textit{Ans.}$$

(*c*) In this case, the load would be 10^4 (both angles = 0)

$\therefore$ Total impedance $= 2 \times 10^4$

$$\therefore \qquad I = \frac{10^2}{2 \times 10^4} = \frac{1}{200}$$

$$\therefore \qquad P = I^2R = \frac{1}{4 \times 10^4} \times 10^4$$
$$= 250 \text{ mW} \quad \textit{Ans.}$$

This shows that by using a transformer, the transferred power in this particular case is increased more than 100 times.

Note that satisfying the second condition produced a further power increase of only about 2 per cent.

Transformers with more than two windings

The same theory applies as for ordinary transformers—*i.e.*, voltages are proportional to turns, and the total effect of all currents on the flux is zero.

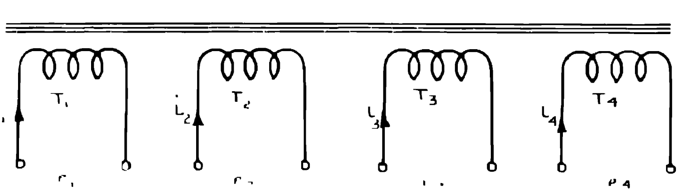


FIG. 205—Transformer with more than two windings.

Consider Fig. 205, which shows a transformer with four windings. The equations are :—

The transformer voltage equation :—

$$\frac{e_1}{T_1} = \frac{e_2}{T_2} = \frac{e_3}{T_3} = \frac{e_4}{T_4} \tag{19}$$

and *the transformer current equation* :—

$$i_1T_1 + i_2T_2 + i_3T_3 + i_4T_4 = 0 \tag{20}$$

It is not usually possible to introduce the idea of transferred impedances, and problems are best tackled from first principles and the application of Kirchhoff's law.

Example.—A three-winding transformer having turns t, t, and $n \cdot t$, is connected, as in Fig. 206, to three impedances R, S, and Q, and to a generator A that produces a current I. Find the current x flowing through R.

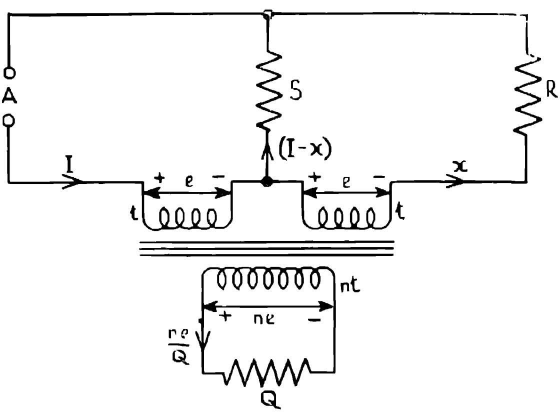


FIG 206 – Example of three-winding transformer.

Let the voltage appearing across one of the transformer windings with t turns be e; the voltages across the other windings can then be written in at once by applying the transformer voltage equation (equation 19), since the turns ratios are known. Note that instantaneous polarities for the voltages, as for the currents, have been inserted.

The current through the impedance Q can be found at once from the voltage across it: this current is $\dfrac{ne}{Q}$, as shown in the figure. The transformer current equation (equation 20) then becomes:—

$$\frac{ne}{Q} \cdot nt = I \cdot t + x \cdot t$$

$$\therefore \qquad e = I \cdot \frac{Q}{n^2} + x \cdot \frac{Q}{n^2} \qquad (i)$$

Applying Kirchhoff's law to the right-hand mesh (*i.e.* that including S and R),

$$e = S\,(I - x) - R \cdot x$$

$$\therefore \qquad e = I \cdot S - x \cdot (S + R) \qquad (ii)$$

Solving for x by subtracting equation (i) from equation (ii):—

$$0 = I\left(S - \frac{Q}{n^2}\right) - x\left(S + R + \frac{Q}{n^2}\right)$$

$$\therefore \qquad x = I \, . \frac{\left(S - \dfrac{Q}{n^2}\right)}{\left(S + R + \dfrac{Q}{n^2}\right)}$$

Note that $x = 0$ if $\dfrac{Q}{n^2} = S$.

LOSSES AND EQUIVALENT CIRCUITS

Up to this point it has been assumed that the transformer was perfect, but in practice there will be losses. The various types of loss and their effect will now be considered; the most convenient way of doing this is by drawing an equivalent circuit.

The losses are divided into four groups: Iron losses (in the core), copper losses (in the windings), flux leakage losses, and self-capacity.

Iron losses

(*a*) *Magnetising current*.—If the inductance of the primary is not infinite, current will flow in the primary when the secondary is off load; it is this current that produces the flux Φ. As Φ is constant for all loads, this current will also be constant, and must be added to the total current on load. It will be 90° behind the primary voltage E_1, and is therefore represented in the equivalent circuit as being caused by an inductance L_p across the primary (*see* Fig. 207*a*). This is, of course, the inductance measured across the primary with the secondary off load, and should be as large as possible. It is drawn on the equivalent circuit across one side of the transformer only. As it is a reactance, it does not introduce a *power* loss.

(*b*) *Eddy current loss*.—The alternating flux, as well as producing voltages in the windings, produces voltages in the metal of the core, causing eddy currents to circulate. As the core material has resistance, this effect is equivalent to a small extra resistive load on the transformer, and it is constant with the flux. This loss is reduced by using insulated laminations for the core, thereby giving the core a very high resistance to eddy currents.

(*c*) *Hysteresis loss*.- Due to hysteresis, losses occur in the process of magnetising the core. These are *power* losses which appear in the form of heat. As power can be represented as occurring in a resistance and not in a reactance, this loss is equivalent to a small extra resistive load, which is constant with flux.

Eddy current and hysteresis losses combine to form a single resistive load on the supply. As this load is constant with flux, and therefore with applied primary voltage, it will be represented by a resistance R_0 across the primary (*see* Fig. 207*a*); and as the loss is small, this resistance will be large.

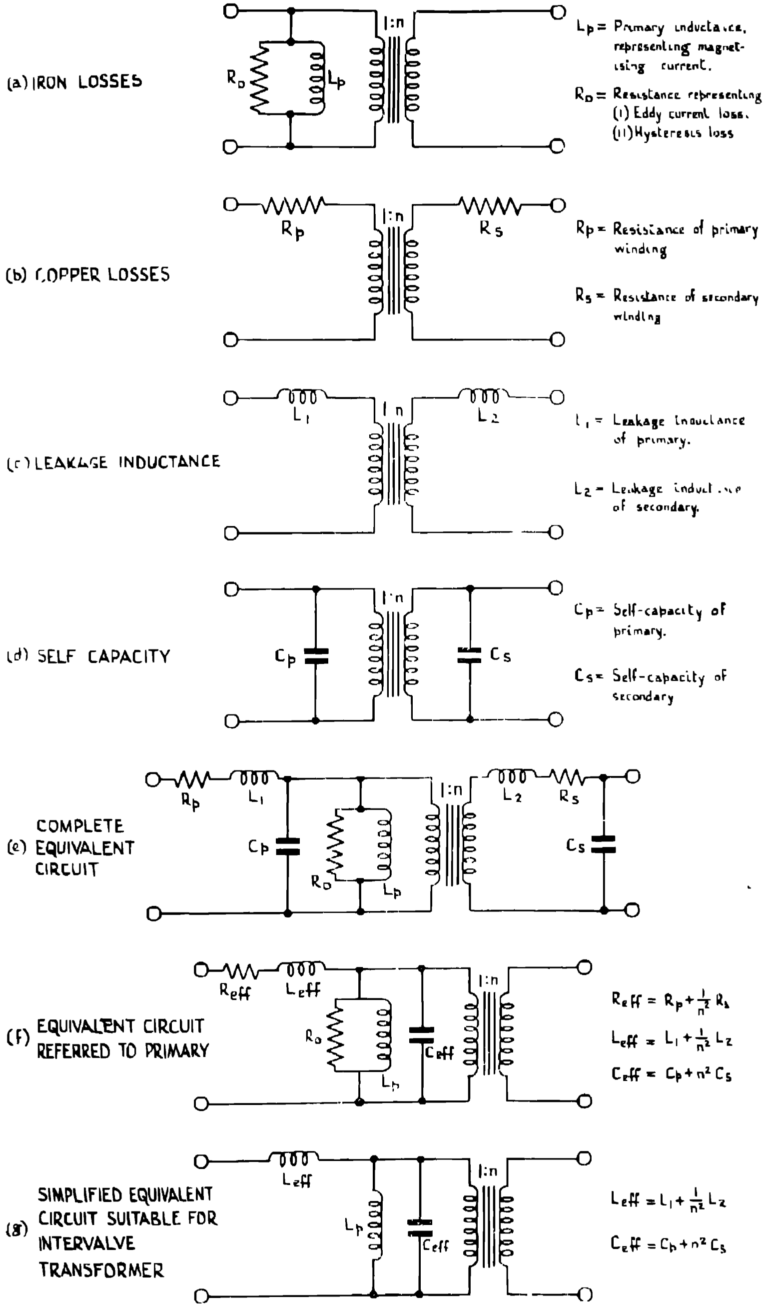


FIG. 207.—Equivalent circuits for transformers having losses.

Copper losses

These are due simply to the DC resistance of the windings, and are represented by the resistances R_p and R_s in Fig. 207*b*. The loss due to these is an " I^2R " loss, which increases with the square of the load current.

Flux leakage losses

It has been assumed that all the flux produced by the primary cuts the secondary turns ; in practice a small amount will not. It will however cut the primary turns, and so produce self-inductance. The effect of flux leakage in the primary can be represented by a small series inductance L_1, as shown in Fig. 207*c*. There will similarly be flux leakage between the secondary and the primary represented by L_2.

Self-capacity of windings

The presence of internal winding capacities may have to be taken into consideration. These are of no importance at power frequencies, but they have a large effect on the behaviour of the transformer at audio frequencies. They can be represented by condensers C_p and C_s across primary and secondary, as in Fig. 207*d*.

Complete equivalent circuit

The equivalent circuit for a transformer, including all the above losses, is therefore as shown in Fig. 207*e*.

It is often convenient to simplify the circuit of Fig. 207*e* by transferring L_2, R_S and C_s to the primary, where they become $\frac{T_1^2}{T_2^2} L_2$, $\frac{T_1^2}{T_2^2} R_S$, and $\frac{T_2^2}{T_1^2} C_S$. They can then be combined with L_1, R_p and C_p to give what are known as the effective leakage inductance, resistance and capacity, referred to the primary—that is, L_{eff}, R_{eff}, and C_{eff} in Fig. 207*f*. Alternatively, L_1, R_P, and C_P could, if desired, be transferred to the secondary.

The complete equivalent circuit can be further simplified by neglecting any losses that may be small. Thus in the case of transformers used in valve amplifiers, the resistance of the windings is usually very small, as also are the eddy current and hysteresis losses. In such cases, R_P, R_S and R_0 can be omitted, and the equivalent circuit, referred to the primary, is then as shown in Fig. 207*g*.

TRANSFORMERS WITH SMALL COUPLING

A transformer can be regarded as two windings possessing self and mutual inductance ; this is indeed the most satisfactory way of dealing with transformers in which the coupling is small.

It has been seen in Chapter 3 (page 170) that if two circuits have mutual inductance M, a current i_1 in the primary will induce

a voltage $\frac{Mdi_1}{dt}$ in the secondary. If there is also a current i_2 flowing in the secondary, it will produce a voltage $-L_s\frac{di_2}{dt}$ in the secondary due to self-inductance. If both i_1 and i_2 are flowing, their effects on flux are additive, so their induced voltages are additive. Hence the total induced voltage in the secondary is :—

$$e_2 = M\frac{di_1}{dt} - L_s\frac{di_2}{dt} \tag{21}$$

Similarly,

$$e_1 = M\frac{di_2}{dt} - L_P\frac{di_1}{dt} \tag{22}$$

These equations apply for any waveform of current. Dealing with pure sine-waves only, and using the vector representation, it has already been seen that $-L_P\frac{di_1}{dt}$ becomes $-j\omega L_P i_1$; similarly $M\frac{di_1}{dt}$ becomes $j\omega M i_1$. Hence the equations for induced voltages become :—

$$e_2 = j\omega M i_1 - j\omega L_S i_2 \tag{23}$$

$$e_1 = j\omega M i_2 - j\omega L_P i_1 \tag{24}$$

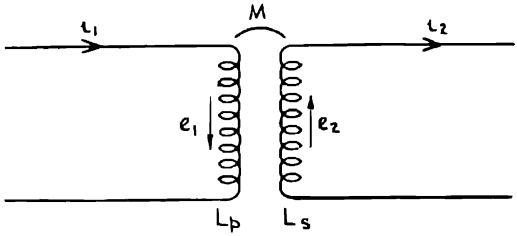


FIG. 208.—Transformer represented by two coupled circuits.

Fig. 208 shows the sense in which these induced voltages are measured.

Suppose now an external voltage E is applied to the primary, and an impedance Z is connected across the secondary. The effective primary impedance will now be calculated (*see* Fig. 209).

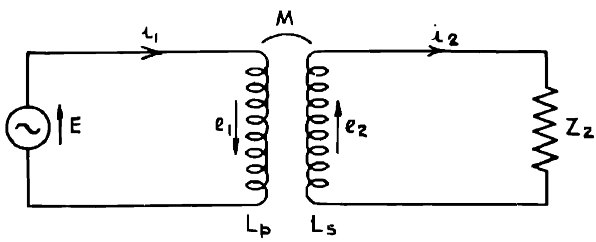


FIG. 209.—Transformer of Fig. 208 connected between generator and load.

The equations 23 and 24 hold. Apply Kirchhoff's Law to the primary :—

$$E + e_1 = 0$$

$$\therefore \quad E = -e_1 = -j\omega M i_2 + j\omega L_P i_1 \text{ (from 24)}$$

$$\therefore \quad E = j\omega L_P i_1 - j\omega M i_2 \tag{25}$$

Apply Kirchhoff's Law to the secondary :—

$$e_2 = i_2 Z_2$$

$$\therefore \quad j\omega M i_1 - j\omega L_S i_2 = i_2 Z_2$$

$$\therefore \quad j\omega M i_1 = i_2 (Z_2 + j\omega L_S) = i_2 Z_S$$

$$\therefore \quad i_2 = \frac{j\omega M i_1}{Z_S} \tag{26}$$

where $Z_S = Z_2 + j\omega L_S$ = total secondary impedance.

Eliminating i_2 from 25 and 26 gives :—

$$E = j\omega L_P i_1 - j\omega M \left(\frac{j\omega M i_1}{Z_S}\right)$$

$$\therefore \quad E = j\omega L_P i_1 + \frac{\omega^2 M^2}{Z_S} i_1$$

$$\text{Primary impedance } Z_P = \frac{e}{i_1} = j\omega L_P + \frac{\omega^2 M^2}{Z_S} \tag{27}$$

The second term is known as the " reflected impedance " from the secondary.

It can be shown that the result given in equation 27 reduces to $Z_P = \frac{T_1^2}{T_2^2} Z_2$ if the transformer is perfect, a result already obtained on page 248. For, if the transformer is perfect, L_P and L_S tend to infinity. $M^2 \rightarrow L_P L_S \rightarrow \infty$, and $\frac{L_P}{L_S} \rightarrow \frac{T_1^2}{T_2^2}$

Equation 27 gives :—

$$Z_P = j\omega L_P + \frac{\omega^2 M^2}{Z_S}$$

$$= j\omega L_P + \frac{\omega^2 M^2}{j\omega L_S + Z_2}$$

$$= \frac{-\omega^2 L_P L_S + \omega^2 M^2 + j\omega L_P Z_2}{j\omega L_S + Z_2} \tag{28}$$

But $\quad L_P L_S = M^2$

$$\therefore \quad Z_P = \frac{j\omega L_P Z_2}{j\omega L_S + Z_2} \tag{29}$$

As $L_S \rightarrow \infty$, Z_2 can be neglected in the denominator.

$$\text{Hence} \quad Z_P = \frac{L_P}{L_S} Z_2 = \frac{T_1^2}{T_2^2} \cdot Z_2 \tag{30}$$

Equation 17 is thus verified.

MAINS TRANSFORMERS

The core

As mains transformers deal only with a low-frequency supply, usually 50 c/s, iron cores may be used. These are laminated to reduce eddy currents and related losses, each lamination being insulated by a coating of shellac or similar substance. The shape is normally as shown in Fig. 210, and suitable proportions are

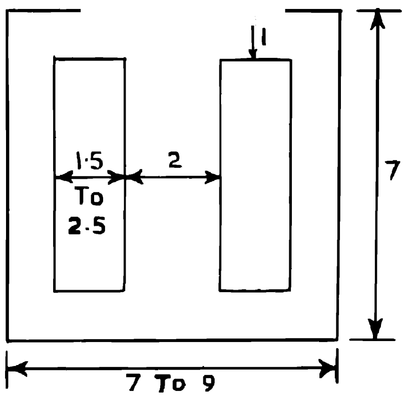


FIG. 210—Typical transformer core stamping

indicated. The core material is usually a silicon-steel alloy, such as stalloy, which has a high permeability and high resistivity, and consequently low losses.

The stack, or thickness, should be between 1 and $1\frac{1}{2}$ times the width of the centre limb. This permits ease of winding and is not thin enough to depreciate the transformer efficiency.

The optimum effective core area—*i.e.* the cross sectional area of the centre limb—is given approximately by:

$$A = \frac{\sqrt{W}}{5 \cdot 58} \text{ sq. in.}$$

where W = volt-amperes output.

The windings

The windings are usually of insulated copper wire, and are made around the centre limb of the core, the other limbs serving to complete the magnetic circuit and so reduce its reluctance. The normal form of insulation for the wire is enamel, and the layers are interleaved with paper. To simplify construction, however, silk or cotton covered wire may be used and in this case interleaving will not be necessary. Some thicker insulation should be provided between the separate windings and between the core and the first layer.

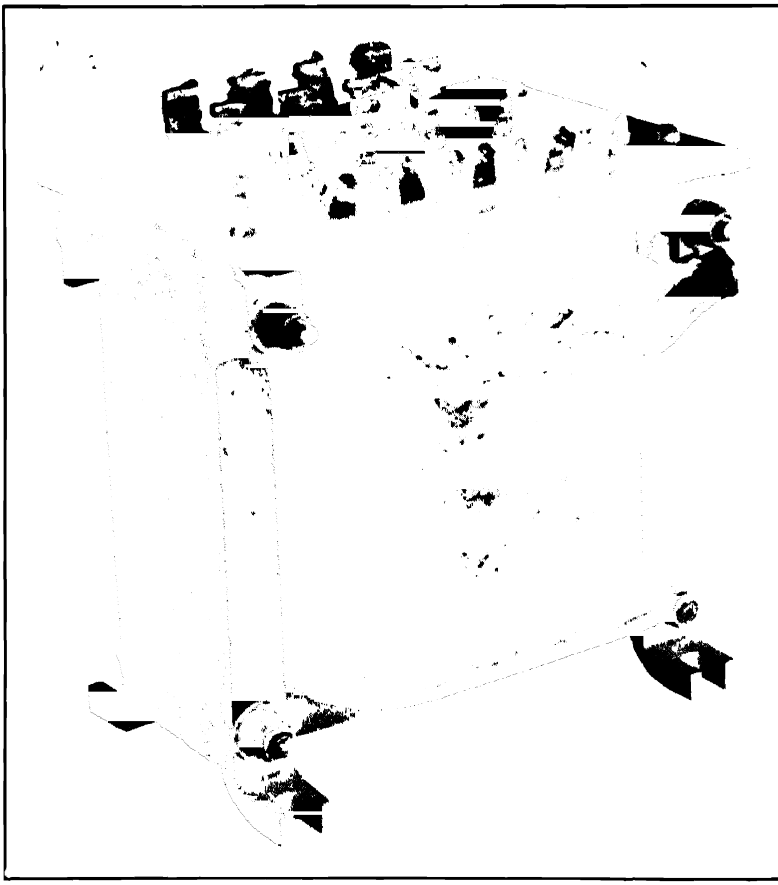


PLATE 9.—Mains transformer.

The number of turns required on the primary is given by the formula

$$N = \frac{0{\cdot}225 \times 10^8 E}{f B_{max} A}$$

where E = primary voltage
f = frequency in c/s
B_{max} = maximum AC flux density
and A = cross-sectional area of core in sq. inches.

Power loss and temperature rise

It should be noted that with mains transformers, considerable power may be transferred; as the efficiency of an average transformer is about 90 per cent., the power wasted in the transformer and converted to heat energy may be quite large. This wasted energy is reduced to a minimum by the choice of a suitable gauge of wire for the primary and secondary windings; a very thin wire would give too great a resistance and cause too great a heat loss, with excessive rise in temperature and disastrous results. The standard upon which choice of wire should be based is that the gauge of wire for both windings should have a cross-sectional area of 1200 to 1500 circular mils per ampere. The following table has been drawn up to enable the correct gauge to be found at a glance :—

TABLE XI

Current-carrying capacity of copper wires

Current in Amperes	Required area in circular mils	Appropriate Wire Size S.W.G.	Ohms per 1000 feet at 60° F.
0·001	1·2	49	7,077
0·01	12·0	43	786
0·1	120·0	31	76
0·5	600·0	22	13·0
1·0	1200·0	20	7·9
2·0	2400	17	3·25
3·0	3600	16	2·49
4·0	4800	15	1·97
5·0	6000	14	1·59
6·0	7200	13	1·20
7·0	8400	13	1·20
8·0	9600	12	0·94
9·0	10,800	12	0·94
10·0	12,000	11	0·76
12·0	14,400	10	0·62
15·0	18,000	9	0·49
20·0	24,000	8	0·40
25·0	30,000	7	0·33

Note.—A circular mil is the area of a circle of diameter 0·001 inch.

A check that the temperature rise will not be excessive can be obtained by adding the copper and iron losses and dividing the sum by the total surface area of the transformer. If the loss per square inch is less than 0·5 watts, then the temperature rise on load should not exceed 40° C. and operation will be satisfactory.

AUDIO FREQUENCY TRANSFORMERS

Line transformers

Line transformers are used to match lines, say, to 600 ohms for connecting to exchanges, *etc.* The primary inductance L_p must be large compared with 600 ohms at all frequencies transmitted, to prevent it causing a shunt loss in a 600 ohms circuit ; it is therefore usually of the order of 1H. As power is to be transmitted, the winding resistances must be small ; generally several hundred turns of medium gauge wire are used on a stalloy core with no air-gap. This gives a sufficiently large inductance, and a resistance on the 600-ohm side of less than 200 ohms. The disadvantage of this is the low saturation-level for the core, which prevents the transformer from being used with any large polarising DC or with 17 c/s AC ringing. If the latter is required, more turns must be used on a large core.

LETTER	IMPEDANCE RATIO LINE		EQUIPMENT
A	1	:	1
B	1·6	:	1
C	2·6	:	1
D	0·62	:	1
E	0·38	:	1
F	0·286	:	1
G	0·133	:	1
H	2	:	1
I	0·5	:	1

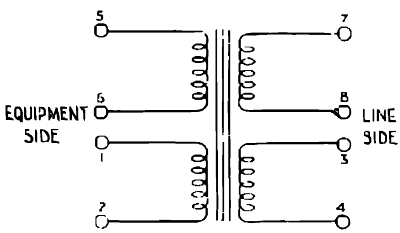
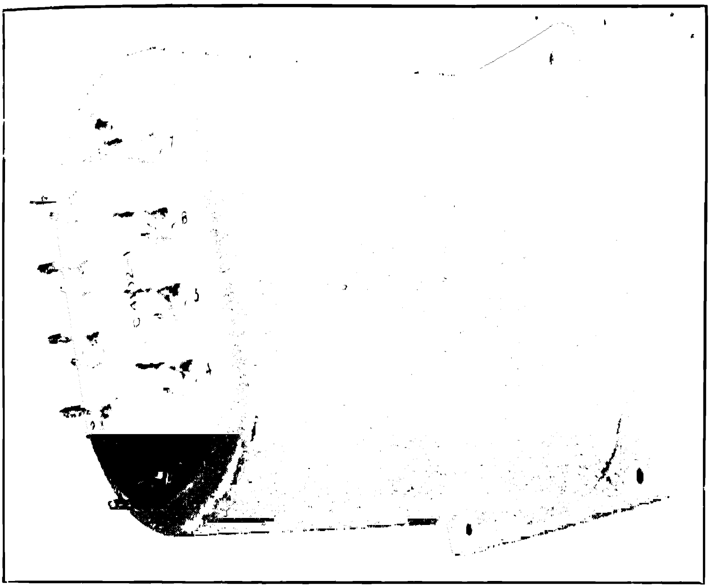


FIG 211 – Line transformers, Types 48 and 50.

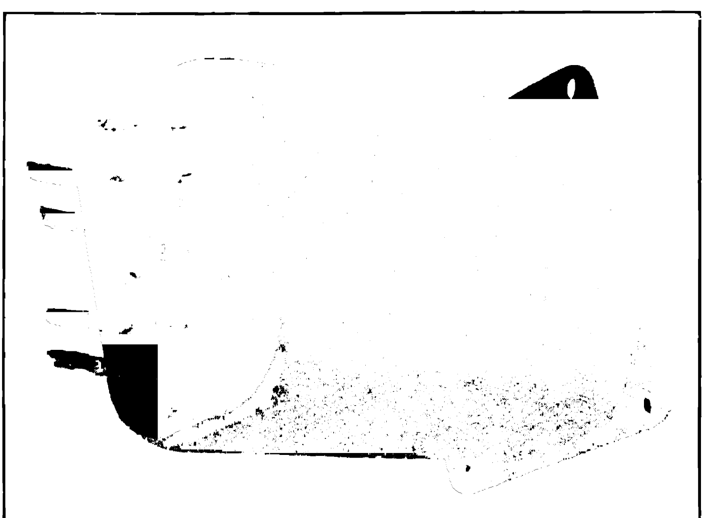
Line transformers are usually of the P.O. type 48 or 50. The type 48 will not pass 17 c/s ringing, whereas the type 50 will. The *impedance* ratio is indicated by a suffix in accordance with Fig. 211.

Amplifier input transformers

This covers transformers used for amplifiers, *etc.*, which work with a very large impedance across the secondary (such as the grid circuit of a valve). The primary impedance must be high in most cases (about 30H at audio frequencies if a very large input impedance is required), but as there is usually no polarising DC, this can be obtained by using a core with a large permeability—*e.g.* " radiometal " or " mumetal "—with several thousand turns. If a large step-up ratio is required, the number of turns on the secondary must be very large. The secondary resistance is not important, as no current is flowing. The limiting factors are the secondary self-capacity referred to the primary, and leakage inductance, both of which increase rapidly with the number of turns. This gives a limit of about 1 : 5 if a large input impedance is desired ; if the primary inductance can be dropped to, say, 1H, ratios up to 1 : 20



Type 50



Type 48

PLATE 10.—Line transformer

can be obtained. The secondary capacity will resonate at some frequency with the leakage inductance, and if the secondary capacity is large, this resonance will occur within the frequency range of the amplifier, causing a peak in the response. This usually happens at about 10 kc/s.

Fig. 212 shows the equivalent circuit; in (*b*) the irrelevant components have been omitted. It will be seen that when L and C resonate, the voltage across C will be large, producing a peak in

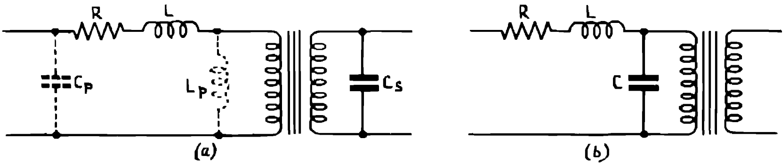


Fig. 212.—Equivalent circuit of input transformer.

the response. This can be damped by increasing (artificially) the resistance R, either by a series resistance in the primary or a resistance across the secondary. Above this resonant frequency the response drops, due to the shunting effect of C.

Intervalve transformers

These are similar to input transformers except that DC usually flows through the primary, and the primary inductance *must* be large as it represents the anode load on the first stage (*see* p. 390). This requires a core with an air-gap to prevent saturation, and a larger number of turns on the primary. This reduces the turns ratio to a maximum of about 1 : 3½.

Amplifier output transformers

These usually carry DC in the primary, and the primary inductance must be large compared with the load reflected from the secondary. As power is being transferred, the resistance of the windings must be low. The primary usually has several thousand turns of fine copper wire (30–40 S.W.G.), and the secondary is determined by the output impedance. The core is of stalloy or radiometal, with an air gap.

Auto-transformers

An auto-transformer is one in which the primary is part of the secondary, or vice versa (*see* Fig. 213).

The same formulae for voltage and current ratio still apply, *i.e.* $\frac{E_1}{E_2} = \frac{T_1}{T_2} = \frac{I_2}{I_1}$.

Its main advantage is the fact that losses are reduced. For the current flowing through the secondary is, from Fig. 213, $(I_2 - I_1)$. This is less than I_2, and hence causes smaller copper and flux leakage losses. This is particularly noticeable when $\frac{T_1}{T_2} \simeq 1$, when

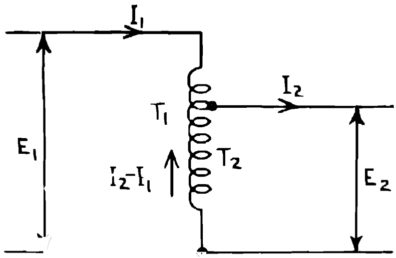


FIG 213 —Auto transformer

I_1 and I_2 will be almost equal and the losses very small. Hence auto-transformers are most useful when a small turns ratio is required.

Fig 214 shows a particularly useful form of auto-transformer used when it is desired to change the impedance of a circuit, and

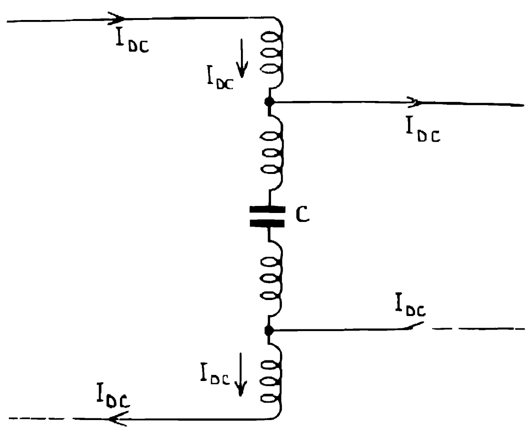


FIG. 214 —Auto-transformer with blocking condenser at centre tap.

yet still provide a DC path say, for signalling or DC testing. The condenser C prevents a DC short-circuit being placed across the circuit by the transformer winding

Core materials

1. *Iron.*—The maximum permeability depends on the purity, and is of the order of 10,000 to 20,000. Saturation level is high. Hysteresis and eddy current losses are high.

2. *Silicon-iron alloys.*—Losses smaller than iron. Stalloy (4 per cent. silicon) is the most common type. Maximum permeability is about 15,000.

3. *Nickel-iron alloys.*—These alloys are characterised by low saturation level and high permeability at low flux densities.

Permalloy.—78 per cent. nickel, permeability up to 100,000.

Hypernik.—50 per cent nickel, similar to permalloy, but losses are smaller and saturation level is higher.

Mumetal.—73 per cent. nickel, 22 per cent. iron, 5 per cent. copper. High permeability, low hysteresis and eddy-current losses, low saturation level.

Radiometal.—Similar to mumetal, but with lower eddy-current losses and lower initial permeability.

Perminvar. 45 per cent. nickel, 30 per cent. iron, 25 per cent. copper. Constant permeability at low fluxes. Negligible hysteresis loss. Loses characteristics if not carefully treated.

4. *Cobalt alloys.*

Permendur.—50 per cent. iron, 50 per cent. cobalt. Has high permeability at high flux densities.

Pure iron is seldom used in transformers. Silicon-iron is used mainly in mains transformers and power transformers at audio frequencies, and also for smoothing chokes, etc. It is also used for interstage transformers carrying a polarising DC current.

Nickel-iron alloys are used principally for audio frequency transformers working at low voltages (such as found in telephony). Their high initial permeability enables large inductance values to be obtained without an excessive number of turns.

Cobalt alloys. Permendur is used chiefly for telephone receivers, etc., where the flux density is high.

At high carrier frequencies, thin nickel-iron stampings are generally used. Where losses must be kept to a minimum, toroidal dust cores can be used.

Distortion due to B-H curve

If a sinusoidal voltage is applied to a transformer, it follows from the basic theory that the flux must be sinusoidal. If no saturation takes place in the core, the B-H curve is linear, and the magnetising current will also be sinusoidal. If, however, saturation does occur, to maintain a sinusoidal flux the magnetising current cannot also be sinusoidal, but must increase when saturation occurs at each half cycle. A typical waveform is illustrated in Fig. 215.

This clearly causes distortion. If no DC is present, only *odd* harmonics will be produced. If two frequencies are applied simultaneously, this will lead to cross-modulation. As both these effects are undesirable, it is most important that hysteresis and saturation should be reduced to a minimum in audio frequency transformers.

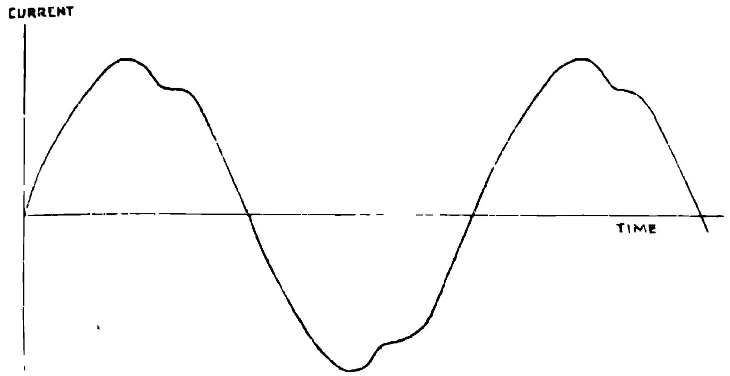


FIG. 215.—Typical waveform of transformer magnetising current for sinusoidal EMF of sufficient amplitude to cause saturation on peaks.

Interwinding capacities

To ensure that two halves of a winding are symmetrical about the centre point, special precautions have to be taken when winding the transformer, for not only must the two halves of the winding have exactly the same number of turns and resistance, but all capacities must also be balanced. The most convenient way of doing this is by winding the two windings together (*i.e.* two wires side by side) and connecting as in Fig. 216.

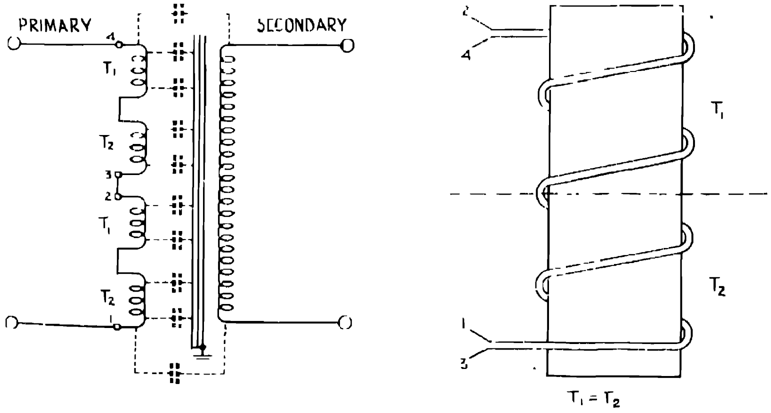


FIG. 216.—Method of balancing interwinding capacities.

It is sometimes also important that the capacity from primary to secondary should be kept to a minimum. This is done by inserting an earthed screen between the windings, increasing their capacity to earth but reducing the capacity between the two. The screen usually consists of a copper sheet covering the winding; the ends should overlap, but must not make electrical contact, since this would give the effect of a short circuited turn.

The principle of the direct current Wheatstone's bridge (see page 129) can be extended to alternating current bridge circuits. Not only are these useful for measuring unknown impedances, but also some forms of AC bridge can be used for measuring frequency, while others find special applications in the circuits of telecommunications equipment.

General case of AC bridge

Let four impedances be connected as in Fig. 217, and an alternating voltage applied between A and B ; then the bridge formed by these four impedances is said to be balanced when the instantaneous PD between C and D is always zero. If the applied voltage

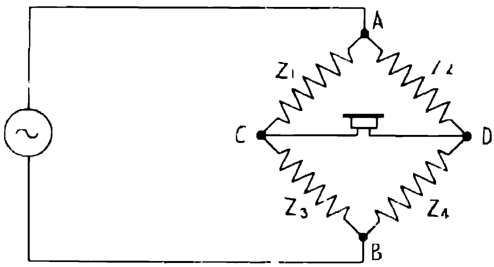


FIG. 217.—General case of AC bridge.

is of audible frequency, the balance condition can conveniently be found by adjusting one or more of the impedances for minimum sound intensity in a telephone receiver connected between C and D.

The condition for balance is found in the same way as for the Wheatstone's bridge, and is

$$Z_1 Z_4 = Z_2 Z_3 \tag{31}$$

This is a vector equation, since Z_1, Z_2, Z_3, and Z_4 are impedances having both modulus and angle. If the impedances are written in the form $|Z|\underline{/\varphi}$, this equation becomes :—

$$(|Z_1|\underline{/\varphi_1}) \cdot (|Z_4|\underline{/\varphi_4}) = |Z_2| \cdot (\underline{/\varphi_2}) \cdot (|Z_3| \underline{/\varphi_3})$$

i.e.
$$|Z_1| \cdot |Z_4| \quad \underline{/\varphi_1 + \varphi_4} = |Z_2| \cdot |Z_3| \quad \underline{/\varphi_3 + \varphi_3} \tag{32}$$

This is equivalent to the two conditions :—

$$|Z_1| \cdot |Z_4| = |Z_2| \cdot |Z_3| \tag{33}$$

and
$$\underline{/\varphi_1 + \varphi_4} = \underline{/\varphi_2 + \varphi_3} \tag{34}$$

Both these conditions must be satisfied for a true balance to be obtained ; two adjustments therefore have to be made, so that the bridge may be balanced with respect to both magnitude and phase angle (or, what amounts to the same thing, with respect to both resistance and reactance).

Impedance bridges

If one of the impedances in a balanced bridge is unknown, its value can be calculated from those of the other three, by equations 33 and 34. Frequently, however, it is desired to know the resistance and the inductance or capacity of a circuit component (e.g. a coil) rather than its impedance at one particular frequency. If the frequency of the supply is known, these can be calculated from the measured impedance.

Bridges that balance at all frequencies

When the equations of balance for certain bridges are solved for the values of the circuit constants (R and L or C) rather than of the impedance the frequency f_0 of the applied EMF vanishes from the equations, and the required values are given in terms of the resistance, inductance, and capacity of the three known arms of the bridge. Bridges in which this occurs are said to be "independent of frequency", they are particularly suitable for measurements of R, L, and C since the frequency of the supply need not be accurately known, and a balance obtained at one frequency will hold at any other (provided that the circuit constants of the four bridge arms do not vary with frequency).

Bridges that balance at only one frequency

A bridge in which the frequency of the applied EMF does not vanish from the balance equations when they are solved for the circuit constants R and L or C is said to be "dependent on frequency". A balance obtained at one frequency on such a bridge will not, in general, hold at any other frequency, and hence if the values of the circuit constants of all four arms of the bridge are known, the frequency of the applied EMF may be calculated. A bridge of this type can therefore be used to measure frequency, in which case it is called a "frequency bridge".

Simplification of general case

In order to simplify the circuit, and to reduce the number of variables in the equations, it is usual in practice to make two arms

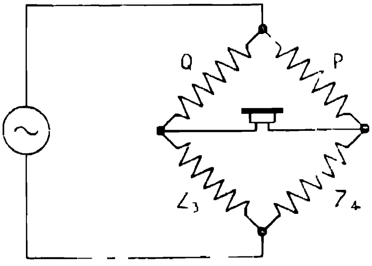


FIG 218—AC bridge with two adjacent impedances replaced by pure resistances.

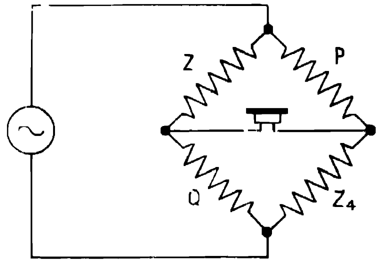


FIG 219—AC bridge with two opposite impedances replaced by pure resistances.

of an AC bridge purely resistive—say, for example, $Z_1 = Q$ and $Z_2 = P$, as in Fig. 218, or $Z_3 = Q$ and $Z_2 = P$, as in Fig. 219. The conditions of balance (equations 33 and 34) thus become :—

For Fig. 218 :-

$$Q \cdot |Z_4| = P \cdot |Z_3|$$

i.e.

$$|Z_4| = \frac{P}{Q} \cdot |Z_3|$$

and

$$\varphi_4 = \varphi_3$$

and for Fig. 219 . -

$$|Z_1| \cdot |Z_4| = P \cdot Q$$

i.e.

$$|Z_4| = \frac{PQ}{|Z_1|}$$

and

$$\varphi_4 = -\varphi_1$$

It can be seen that two conditions are still necessary for a balance to be obtained.

Adjustment of bridges

The two adjustments necessary to balance an AC bridge are not, in general, independent. It is necessary, therefore, to adjust one control until approximately minimum output is obtained from the detector (*e.g.* the telephone receiver in Fig 217), then to adjust the other; and to repeat this process until the output cannot be further reduced.

Certain bridges have been developed in which one control governs the resistive balance, and the other the reactive; while in other bridges, one control enables the modulus of the impedances to be balanced, and the other the phase angle. Even in these bridges the adjustments of the two controls are not completely independent, and the procedure for balancing given above should be followed if accurate results are to be obtained.

Example of an AC bridge balancing at only one frequency (Series resonance bridge)

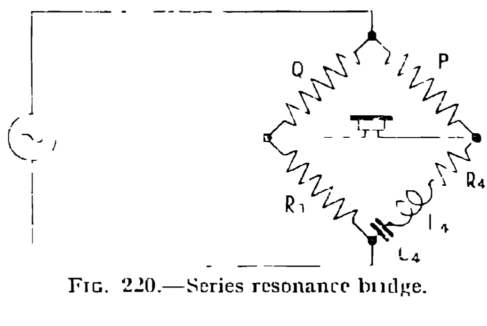


FIG. 220.—Series resonance bridge.

The condition for balance, from equation 32, is :-

$$Q\left[R_4 + j\left(\omega L_4 - \frac{1}{\omega C_4}\right)\right] = PR_3$$

$$\therefore \qquad QR_4 = PR_3 \quad \text{and} \quad \omega^2 L_4 C_4 = 1$$

This bridge can be used to measure an unknown inductive impedance $(R_4 + j\omega \,.\, L_4)$ if P and Q are fixed known resistors, and R_3 and C_4 are variable components whose values are known. Then :--

$$R_4 = \frac{P}{Q} \,.\, R_3 \quad \text{and} \quad L_4 = \frac{1}{\omega^2 C_4}$$

It can also be used to measure an unknown frequency if the value of the inductance L_4 is known. The frequency is then given by :—

$$\omega^2 = \frac{1}{L_4 C_4}, \quad \text{or} \quad f = \frac{1}{2\pi\sqrt{L_4 C_4}}$$

The accurate adjustment of R_3 is necessary to obtain zero output in the telephone receiver, although R_3 does not appear in the formula for frequency.

Example of an AC bridge balancing at all frequencies (Maxwell bridge)

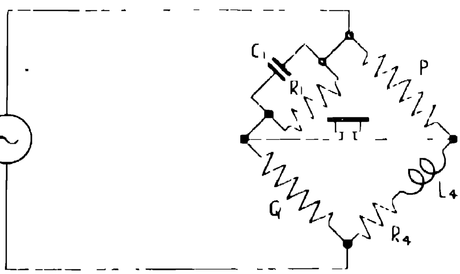


FIG. 221.—Maxwell bridge.

The condition for balance is :—

$$PQ = (R_4 + j\omega L_4) \,.\, \frac{1}{\frac{1}{R_1} + j\omega C_1}$$

i.e. $$\frac{PQ}{R_1} + j\omega C_1 PQ = R_4 + j\omega L_4$$

i.e. $$R_4 = \frac{PQ}{R_1} \quad \text{and} \quad L_4 = PQC_1$$

This bridge also can be used for measuring an unknown inductance, and it has the advantage that the frequency at which the balance is made need not be known accurately. It could, clearly, not be used to measure frequency.

The Wien frequency bridge

An interesting and useful bridge is shown in Fig. 222.

Note that in the lower arms the two resistances and condensers are equal.

The condition for balance is :—

$$\frac{2P}{\frac{1}{R} + j\omega C} = P\left(R + \frac{1}{j\omega C}\right)$$

$$\therefore \quad 2P = P\left(R + \frac{1}{j\omega C}\right)\left(\frac{1}{R} + j\omega C\right)$$

$$\therefore \quad 2P = P\left[1 + 1 + j\left(\omega CR - \frac{1}{\omega CR}\right)\right]$$

$$\therefore \quad \omega CR = \frac{1}{\omega CR}$$

$$\therefore \quad \omega^2 = \frac{1}{C^2R^2}$$

Thus the bridge balances if $f = \dfrac{1}{2\pi CR}$

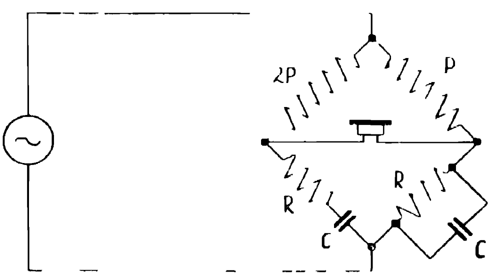


Fig. 222—A.C. bridge (Wien) for measuring unknown audio frequencies

This balances at one frequency only. If both resistances are made to vary simultaneously the bridge can be calibrated to measure frequency. The capacities C are normally fixed, but may be changed simultaneously to another fixed value to alter the frequency range of the bridge.

Summary of frequently used bridges

A table of the most useful impedance bridges is given on pages 272*A* and *B*, together with the equations by which one can calculate the value of the unknown impedance $Z = |Z|\underline{/\varphi} = R + jX = \dfrac{1}{G + jB}$. These equations are obtained by substituting the appropriate values of R_1, Z_1, *etc.*, in equations 33 and 34, and are given in polar co-ordinate form $|Z|\underline{/\varphi}$ in columns (*f*) and (*g*) and also in rectangular component form $R + jX$ in columns (*h*) and (*j*). From the latter, the values of resistance and inductance or capacity that give this impedance are calculated and are given in column (*d*).

In some cases the expressions for R and X are somewhat involved, while those for the corresponding admittance $Y = Z^{-1} = G + jB$ given in columns (k) and (l) are more manageable. In these cases, the impedance Z is more easily represented as a reactance $\left(j\omega L \text{ or } \dfrac{1}{j\omega C}\right)$ *in parallel* with a resistance r (*see* column (c))

The lamp bridge

The lamp bridge is used to indicate when a voltage departs from a preset value, or to maintain a voltage at a preset value. It depends for its action on the fact that the resistance of a lamp increases as the current through it is raised.

The basic circuit is shown in Fig. 223, where R_1, R_2 and R_3 are normal resistances whose values are (for all practical purposes) independent of the current through them, and where R_4 is a lamp.

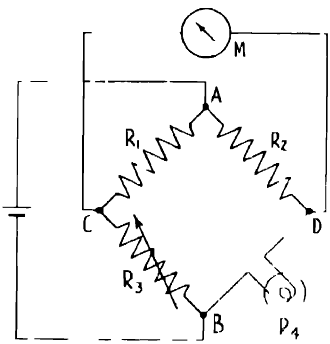


FIG. 223 Lamp bridge calibration on DC from battery

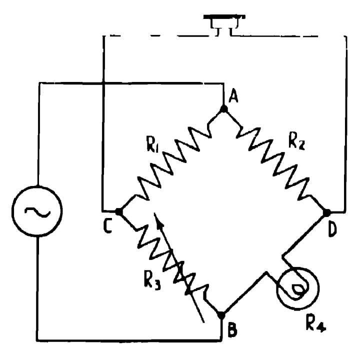


FIG. 224 Lamp bridge calibration on audio frequency

With the standard voltage E applied between terminals A and B, R_3 is adjusted so that the bridge is balanced and the meter reads zero Then $R_1R_4 = R_2R_3$.

If the voltage applied between A and B rises above the standard value E, the current both through R_1 and R_3 and through R_2 and R_4 increases. The increased current through R_4 raises its temperature, and therefore its resistance The bridge then becomes unbalanced, and the potential of D rises above that of C, so that the meter M deflects Similarly, if the applied voltage falls below the standard value E, the resistance of R_4 drops and causes the potential of D to drop below that of C, the meter M then deflects in the opposite direction.

If the applied voltage be alternating, then the output across CD will again be zero when the bridge is balanced; if it be of audible frequency, this condition can conveniently be detected by means of a telephone receiver, as shown in Fig 224. The terminals

A and B can then be connected to any EMF that it is desired to adjust to the same value E; the amplitude of this EMF is then adjusted until no tone is heard in the receiver. This balance is independent of frequency.

When the EMF applied to AB is lower than E, then the voltage appearing across CD is in phase with it; and if it is higher than E, then this voltage is 180° out of phase with it. While this phase change is not evident from the headphones, it can be used automatically to control the output of an oscillator (*see* page 477).

CHAPTER 6

METAL RECTIFIERS AND POWER SUPPLIES

METAL RECTIFIERS

Rectification is the process of converting an alternating current into a direct current. This can be done by the use of the diode valve, which possesses the property of passing current only from the anode to the cathode. A valve, however, is not mechanically strong, and it requires an external power supply for its filament or heater circuit ; as an alternative, there are now two types of metal rectifier, the copper oxide rectifier and the selenium rectifier. Early metal rectifiers were large and cumbersome, but with subsequent development, they are now compact, robust, and efficient, and have become the accepted means of rectification in the majority of smaller power packs.

The copper oxide metal rectifier is made by coating one side of a copper disc with a layer of red cuprous oxide. The layer, being obtained by heat treatment, is very hard. This combination offers a low resistance to current flowing from the oxide to the copper, but an extremely high resistance to current flowing from the metal to the oxide.

The selenium type of metal rectifier is a more recent development. The selenium layer may be formed on almost any type of metal surface, but the one most commonly employed is nickel-plated steel. This layer of selenium is then sprayed with a low melting point tin alloy which forms the " counter-electrode " and makes the assembly mechanically sound.

Considerable research has been carried out to discover exactly where the asymmetrical resistance is introduced. The metals and alloys are all linear resistances and the oxide, though having non-linear resistance, is not asymmetric. The all-important asymmetry of the rectifier is thus assumed to be due to a layer existing between the cuprous oxide and the copper in the first case, and between the selenium and the counter-electrode in the second. This is termed the " barrier-layer " and, though its existence has not yet been definitely proved, it gives a satisfactory explanation of the rectifier action.

As, in practice, the " forward " resistance is not zero and the " backward " resistance is not infinite, considerable heat is evolved during operation, and, if no adequate cooling is provided, the resulting temperature rise will cause decrease of both forward and

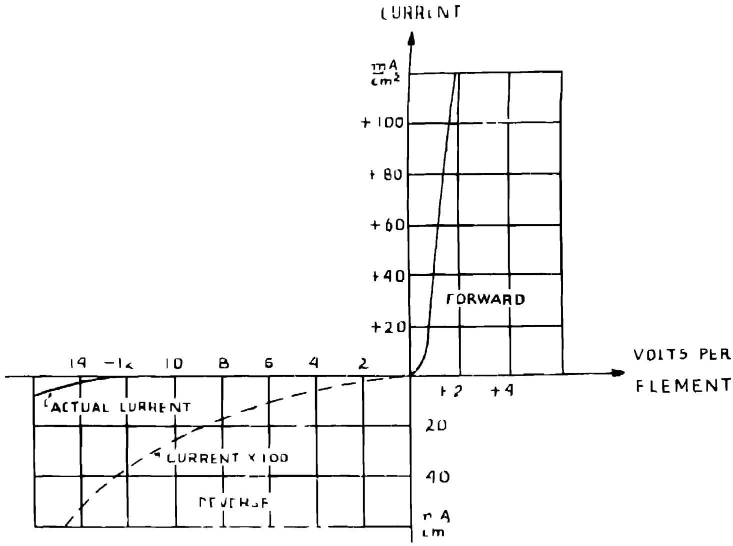


Fig. 225.—Characteristic of current against applied voltage for a typical selenium rectifier

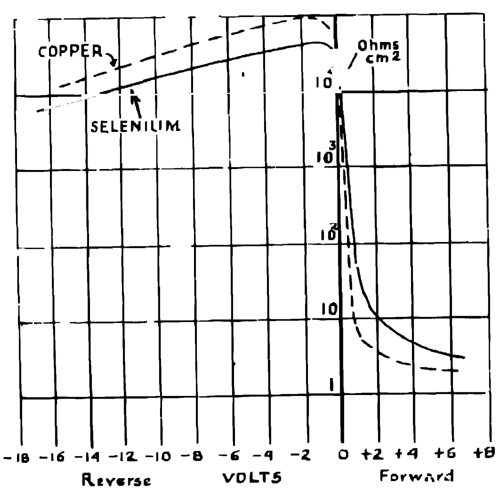


Fig. 226.—Resistance voltage curves

reverse resistances of both types of rectifier, thus altering their performance An excessive rise may cause permanent damage to the rectifier

For these reasons rectifiers are designed for working at a certain ambient temperature usually in the region of 30 C and to stand temperature rise of some 40° C without great depreciation An excessive rise in temperature is frequently prevented by the provision of cooling fins, a free circulation of air round the rectifier is always desirable and in the case of large power plants this air circulation may be improved by the provision of fans

Characteristics of metal rectifiers

It has been stated briefly that the rectifier offers a low impedance to current flowing in the forward direction and a high impedance to current in the reverse direction This property is demonstrated more fully by the characteristic curves

Fig 225 shows the characteristic curve of a typical selenium rectifier The broken curve shows the current (to 100 times the scale) when voltage is applied in the reverse direction It illustrates clearly the very high resistance of the rectifier in this direction, and how it is reduced with increase of voltage At reverse voltages higher than 18 volts the rectifier may be considered to pass current and hence for rectifiers working with higher voltages it will be necessary to use two or more such elements in series

The copper oxide rectifier has a similar characteristic but may be considered to pass current after 8 volts per element in the reverse direction instead of the 18 volts for the selenium element

In the forward direction elements have a high impedance until the applied voltage is above ¼ volt per section for the copper oxide type or ½ volt for the selenium type This property finds application in such apparatus as the acoustic shock absorbers fitted across some telephone receivers

The above current-voltage characteristic can be translated into resistance values, and the resistance voltage curve is thus obtained (see Fig 226)

The copper oxide rectifier

Most of the qualities of this rectifier have been enumerated but little has been said about its actual construction This is illustrated diagrammatically in Figs 227 and 228

In the manufacturing process rings or discs about 1 mm thick, of highly refined copper are heated in air to a temperature just above 1000 C until a layer of cuprous oxide about 0 1 mm thick has formed on the surface The crystal structure of the oxide layer is modified by annealing processes, after which the discs are cooled A thin film of black cupric oxide has by then formed over the red cuprous oxide, and has to be removed by chemical treatment

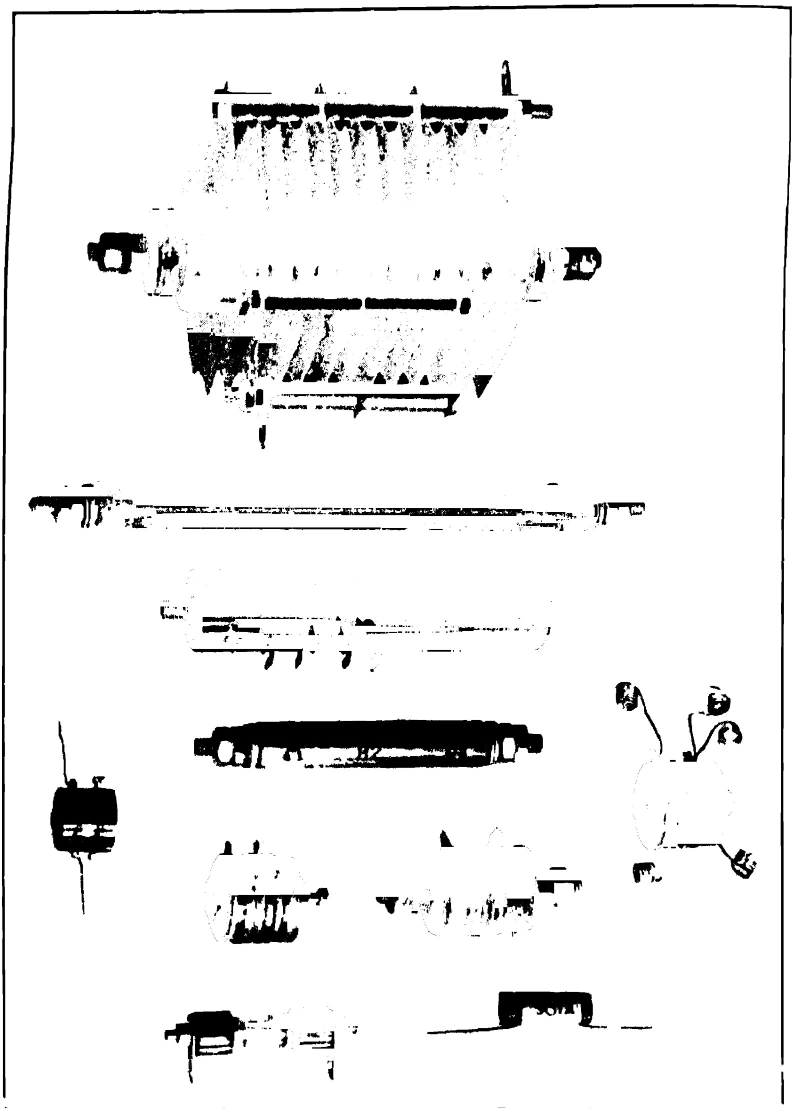


Plate 11 — Copper oxide rectifiers

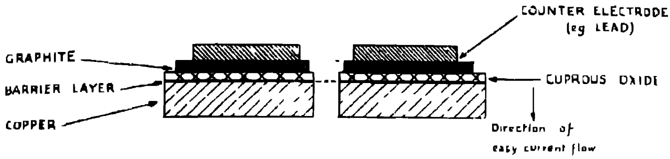


FIG. 227.—Copper-oxide rectifier element (passes conventional current $Cu_2O \rightarrow Cu$).

The cuprous oxide is painted with an aqueous suspension of colloidal graphite and finally covered by the "counter-electrode"—a soft metal (*e.g.* lead) plate or coating.

Copper oxide rectifier elements are rarely used singly owing to the limited reverse voltage that they will stand (6 to 8 v). They are usually connected in series, and must be maintained under

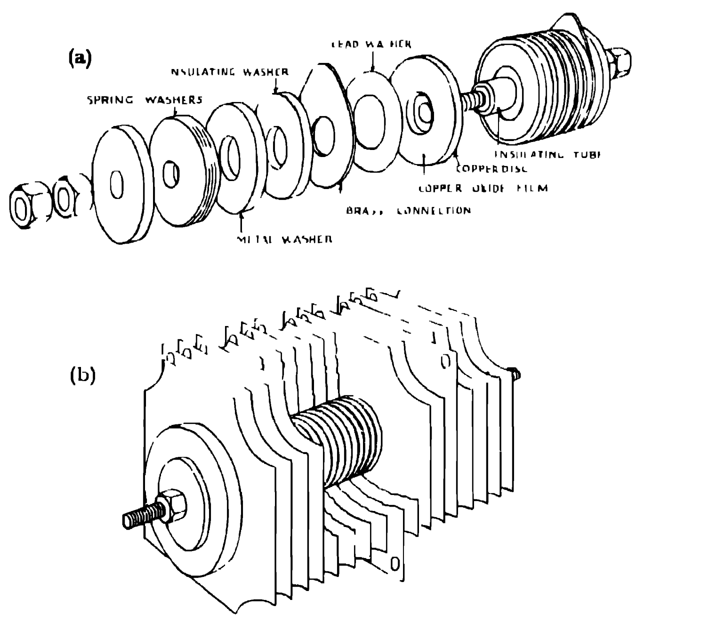


FIG. 228.—Copper-oxide rectifiers (*a*) without cooling fins, (*b*) with cooling fins.

a pressure of 50–60 lb. per square inch to ensure good contact between the lead, graphite and cuprous oxide. This last requirement complicates maintenance, since removal of the pressure tends to change the rectifier characteristics. Thus if one element breaks down, it will probably be found necessary to replace the whole series.

The selenium rectifier

Selenium rectifiers are usually formed on a nickel-plated steel surface Selenium is applied to the base and, by heating in a hot press at about 130° C , is formed into a homogeneous layer about 0 1 mm. thick The temperature is raised to 180–215° C , changing the selenium into a grey crystalline form. A low melting point

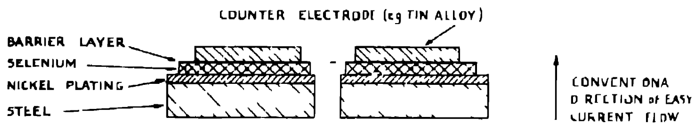


FIG 229 —Selenium rectifier element

alloy is now sprayed on to the selenium layer to act as a counter-electrode, and the manufacture is completed by an electrical forming process which considerably increases the reverse resistance

The selenium rectifier does not require high pressure to ensure contact between the component layers and can thus be easily dismantled and repaired (contrast with the copper oxide rectifier)

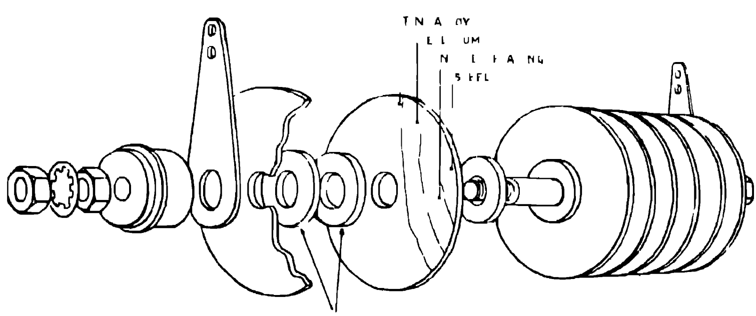


FIG 230 —Selenium rectifier

The lack of pressure enables the separate elements to be spread out, so that the steel plates can act as their own cooling fins.

One peculiar characteristic of the selenium rectifier is that, if it has been employed in the forward direction or with low voltage in the reverse direction, and is then suddenly required to operate with a high reverse voltage, an abnormally high reverse current will flow for an instant , after this, the rectifier resistance increases and the current is cut down to the value normally corresponding to that applied voltage.

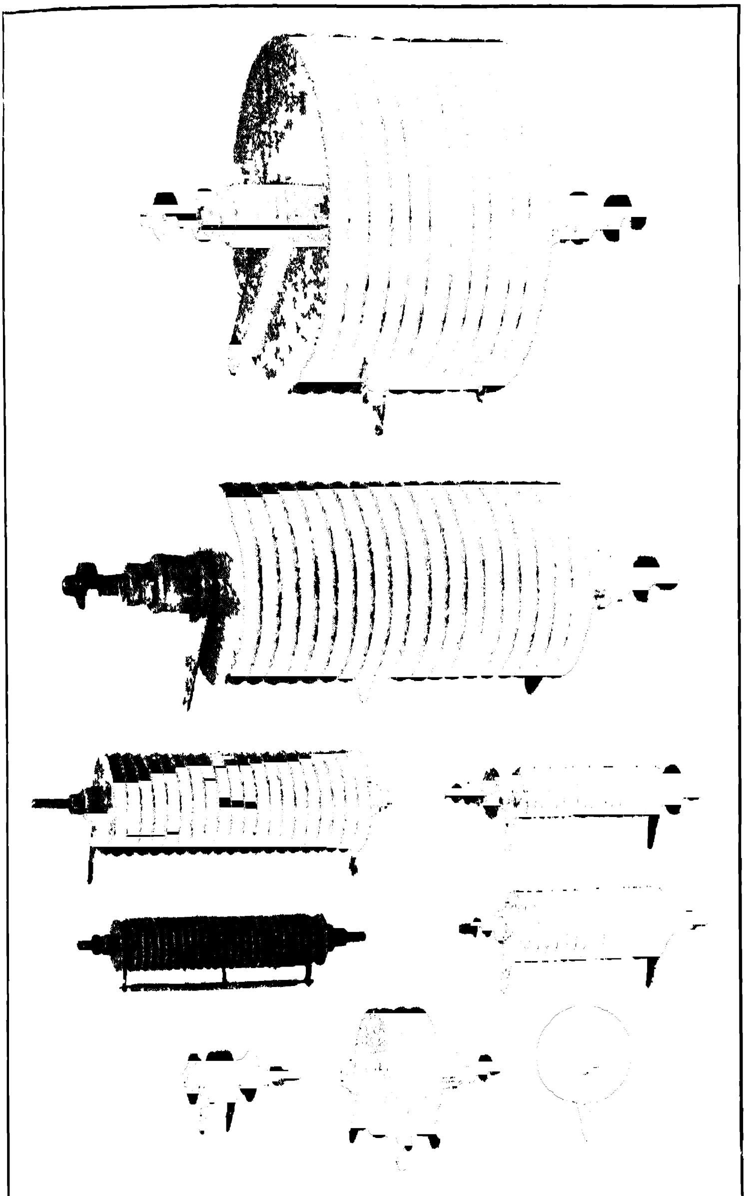


PLATE 12.—Selenium rectifiers.

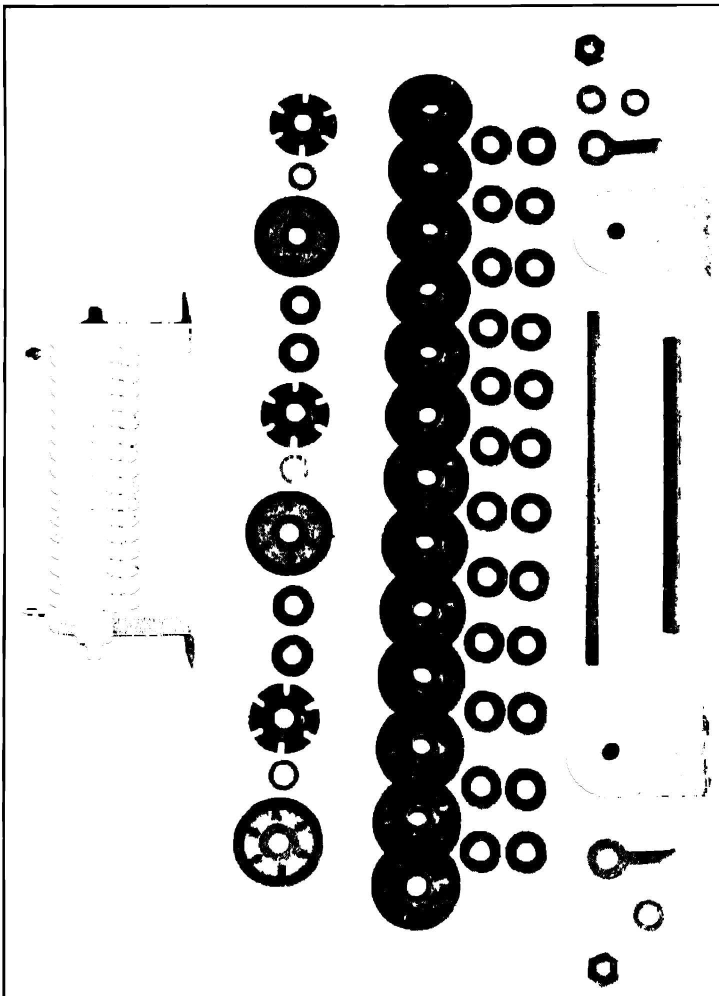


PLATE 17.—Selenium rectifier employing spring spacers showing component parts

Effect of temperature

The forward and backward resistances of both types of rectifier drop with increase in temperature; this is illustrated in Figs 231

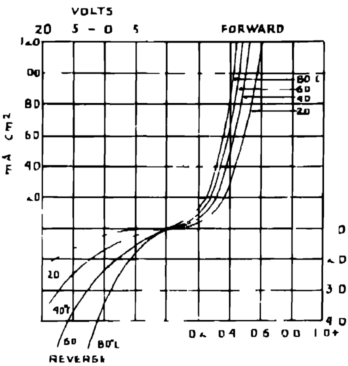


FIG. 231 Effect of temperature on copper oxide rectifier

and 232. This determines the permissible current for a given power dissipation. Selenium rectifiers should be kept below 85° C and copper oxide rectifiers below 55° C. This corresponds to a forward current of ½ to 1 amp per sq in, though this value may be increased to 2–3 amps if special cooling arrangements are made.

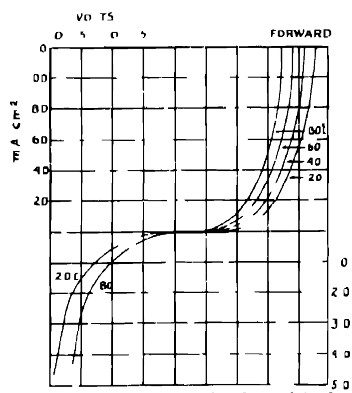


FIG. 232.—Effect of temperature on selenium rectifier

Marking of rectifiers

Two alternative symbols for a rectifier are shown in Fig. 233*a* and *b*. The arrow shows the forward direction of "conventional" current flow; the rectifier resistance is smaller in this direction.

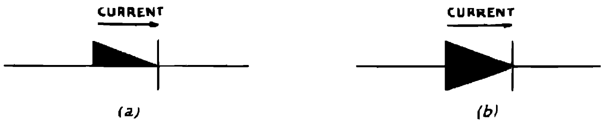


FIG. 233.—Standard symbols for rectifier, showing direction of conventional current flow.

The marking of the rectifier itself is shown in Fig. 234*a* and *b*. At first sight this may seem misleading; it should be remembered that the "+" or RED terminal of a rectifier is the terminal at which the current *leaves* the rectifier.

FIG. 234. Symbols for rectifier, showing standard labelling of terminals.

Self-capacity of rectifiers

All rectifiers possess a certain amount of self-capacity, which is of some importance. For both types of rectifier, it is approximately 0·02 μF per sq. cm. of plate area, and is independent of frequency.

RECTIFICATION CIRCUITS

Half-wave rectification

The simplest method of rectifying an alternating current is to use a half-wave rectifier circuit.

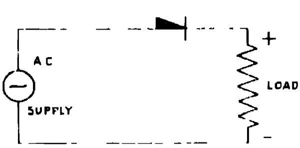


FIG. 235 Series half-wave rectifier circuit.

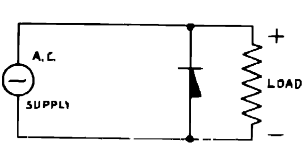


FIG 236 Shunt half-wave rectifier circuit.

Fig. 235 shows a circuit suitable for half-wave rectification; the half-cycle flowing clockwise in the load circuit will pass, while the anti-clockwise half-cycle will be impeded by the reverse resistance of the rectifier. Fig. 236 shows a circuit which will obtain a similar result; but, this time, the rectifier short-circuits the load during the forward half-cycle, and it is the reverse half-cycle which, finding high resistance in the rectifier, passes to the load.

The circuit shown in Fig 235 will operate only if the AC source provides a path for DC. This is not essential for the circuit shown in Fig 236 The load impedance should be less than the reverse impedance of the rectifier in Fig 235 and greater than the forward impedance in Fig 236

In these two systems the output is as shown in Fig 237

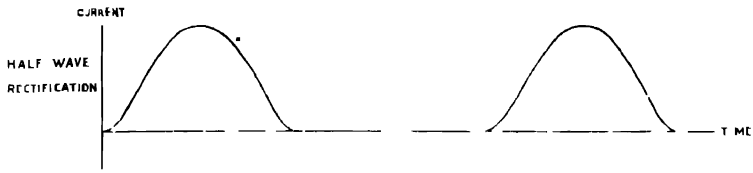


Fig 237 Output from half wave rectifier

This constitutes a wastage of half the voltage supplied and the system is only used when very little power output is required The frequencies contained in the output are the supply frequency and its even harmonics

Full-wave rectification

In other cases full wave rectification is used (*see* Fig 238)

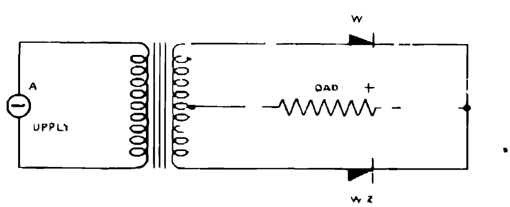


Fig 238 Full wave rectifier circuit

Each half cycle is now utilised On one half-cycle, current will flow through rectifier W_1 to load On the other half-cycle, current will flow via W_2 to the load. The direction of current through the load will be the same for both half-cycles But though "full wave" rectification is now obtained only half of the voltage across the transformer secondary is being applied to the load, and the provision of a centre tap to the transformer is essential The high voltage necessary across the secondary winding of the transformer may be a source of danger

The "bridge" circuit

To avoid these difficulties a rectifier bridge may be used as shown in Fig. 239

The half-cycle of AC flowing clockwise around the circuit finds a low impedance in rectifiers W_2 and W_3, and it therefore

flows through the load from right to left. The next half-cycle, flowing anti-clockwise, finds a low resistance path through rectifiers W_4 and W_1, so that the current through the load is again from right to left.

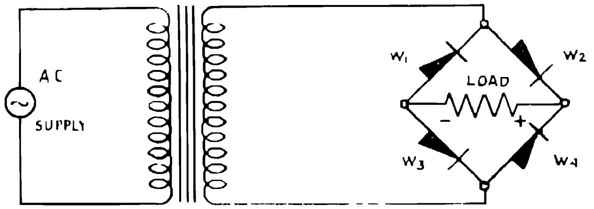


FIG. 239.—Bridge rectifier circuit.

The current reaching the load is now of the form shown in Fig. 240. The output contains even harmonics of the supply frequency, but no fundamental.

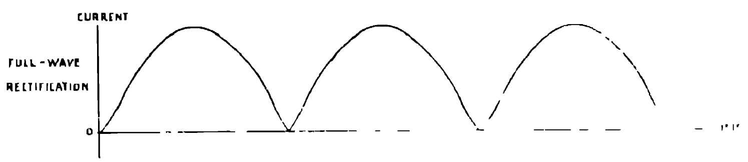


FIG. 240. Output from full-wave rectifier.

This bridge circuit is generally used in cases where currents of 0·5 amp and upwards are required.

Voltage doubler circuits

If a high voltage and small current output is required, a voltage doubler circuit may be employed. Fig. 241 shows a typical full-wave voltage doubler circuit.

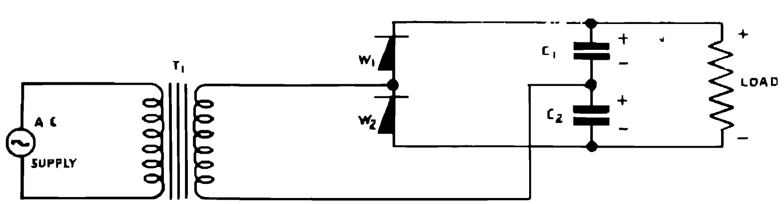


FIG. 241.—Full-wave voltage doubler circuit.

Consider a half-cycle that makes the top of the secondary coil of T_1 positive, and the bottom negative. Current will flow through rectifier W_1, charging the condenser C_1 to the voltage of the supply. During the next half-cycle the condenser C_2 will be charged and, as both condensers are charged as shown, the final voltage across

the load will be twice that of the supply. This will apply, however, only if the load is taking no current; current drain will prevent the condensers from reaching their full charge, and the voltage will in practice never be quite double that of the supply. The percentage by which it falls short of the double value will depend upon the current drain, and the circuit is normally used only when the current drain is very small.

A modified form of the full-wave voltage doubler circuit* is shown in Fig. 242*a*.

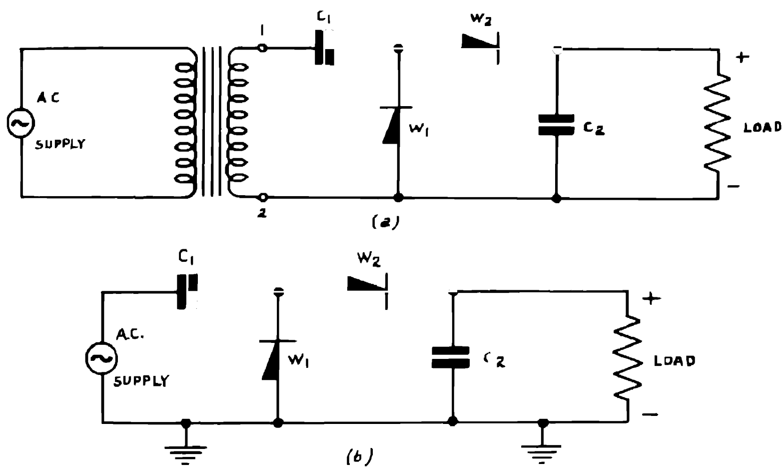


FIG. 242 —Modified full-wave voltage doubler circuit.

Consider a half-cycle that makes terminal 1 of the transformer secondary negative, and 2 positive. Current will flow through rectifier W_1 and charge condenser C_1, but, owing to the low impedance offered by the rectifier W_1 and the high impedance of rectifier W_2, no current reaches the load. During the next half-cycle, rectifiers W_1 and W_2 have high and low impedances respectively. The voltage already across condenser C_1, plus the voltage across the secondary, is therefore applied across condenser C_2 and the load. Thus, though in the first half-cycle no voltage was applied to the load, in the second half-cycle twice the voltage is applied, utilising the charge obtained from the first.

Condenser C_2 acts as a "reservoir"; that is, it stores the charge and maintains the voltage across the load during the half-cycles charging condenser C_1.

The above circuit is often used when one side of the supply and one side of the load are both earthed. In this case the transformer may be dispensed with, as in Fig. 242*b*.

* In American literature, this is often known as a "half-wave" voltage doubler circuit.

Two voltage doubler circuits may be arranged in the form of a voltage quadrupler, as in Fig. 243.

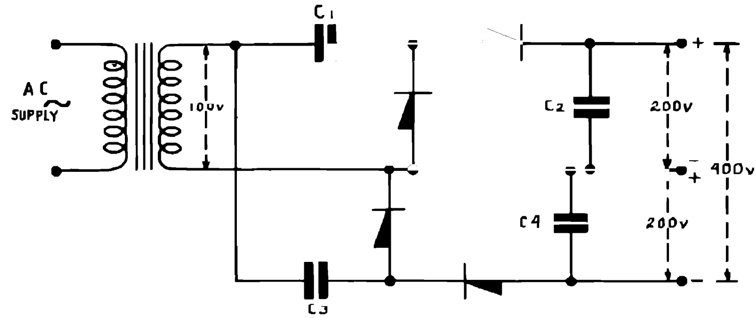


FIG. 243. Full wave voltage quadrupler circuit.

Such circuits are successful only when the current taken is extremely small.

AC METERS

It is clear that meters will be required for making measurements of alternating currents and voltages. Various forms of DC meters have already been described; the effect of AC on these will be considered.

Moving coil meters

The deflection is proportional to the current i; if i is an alternating current of very low frequency the meter will be able to follow the alternations, and will move from side to side about the zero mark. If, however, the frequency is high, the meter will be unable to follow the variations but will read the average value of the current—*i.e.* zero. Hence a pure alternating current will not deflect a moving coil meter, which is therefore of no application in AC measurements. On the other hand, if a current consists of DC plus an alternating component, a moving coil meter in the circuit will read the DC and be unaffected by the AC—a useful property.

Moving iron meters

The deflection of a moving iron meter is proportional to i^2, and for a DC meter the scale is calibrated to read the square root of this, *i.e.* to give a direct reading of current. If i is an alternating current of a sufficiently high frequency the meter will give a steady deflection proportional to the mean value of i^2. This is not zero, but is equal to $[I_{RMS}]^2$, and hence the scale will read I_{RMS}—the root mean square value of the current. Hence moving iron meters will respond to AC, and will give a reading of the RMS value If the current consists of AC and DC, the meter will read $\sqrt{I_{DC}^2 + I_{RMS}^2}$.

The operation of these meters is satisfactory at mains frequencies, but they are seldom used for audio-frequency work. Their principal disadvantage is low sensitivity.

Hot-wire and thermo-couple meters

These both depend upon the heating effect of the current, and hence operate satisfactorily from AC, giving readings of RMS values. Their main advantage is that they can measure AC at any frequency, but they suffer from the disadvantage that they have a very small overload safety margin—the majority of meters of this type being permanently damaged by a 50 per cent. overload.

Rectifier meters

I thas been seen that certain types of meters operate directly from AC; more commonly, however, the AC is rectified and measured on a DC meter. Using metal rectifiers, this method is satisfactory up to frequencies of about 100 kc/s, and enables a

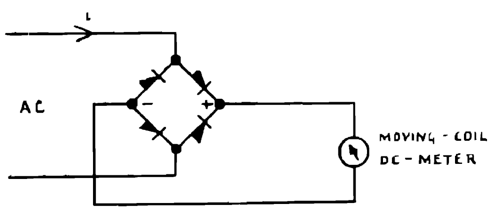


FIG. 244.—Bridge rectifier applied to meter.

moving-coil DC meter to be used with consequent high sensitivity. It is usual to employ a full-wave rectifier bridge for this purpose, consisting of copper oxide elements, the circuit being as shown in Fig. 244.

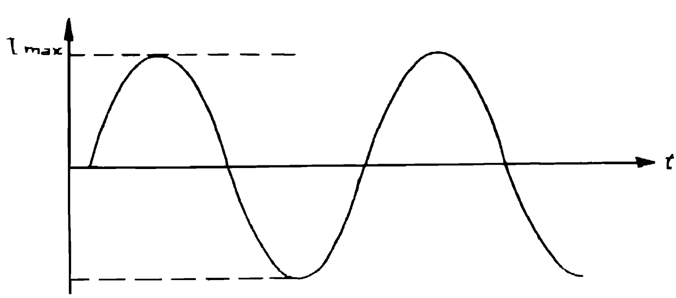


FIG. 245.—Sinusoidal alternating current applied to rectifier.

It is important to know what relationship exists between the alternating current or voltage and the meter deflection. It will be shown that the deflection is directly proportional to the alternating current, but not to the alternating voltage. Suppose the current i is of the form shown in Fig. 245—a sine wave, of peak value I_{max}.

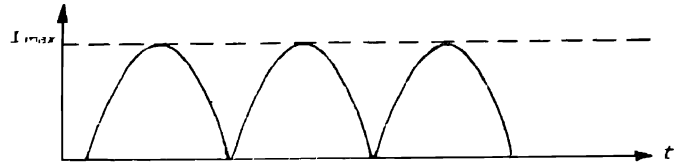


FIG. 246.—Rectified current applied to meter.

The rectified current will be of the form shown in Fig. 246.

The peak value will be I_{max}, for in a metal rectifier working below overload point the leakage current is negligible—less than 1 per cent. The meter deflection will be proportional to the mean value of this waveform; it can be shown that (for a sine wave) this is equal to $0 \cdot 637\ I_{max} = 0 \cdot 9I$, where I = RMS value of alternating current. Hence the reading of a DC meter would have to be multiplied by $1 \cdot 11$ to give the RMS value of the current, or alternatively the scale could be recalibrated. Note that the meter scale will still be linear—*i.e.* the deflection is directly proportional

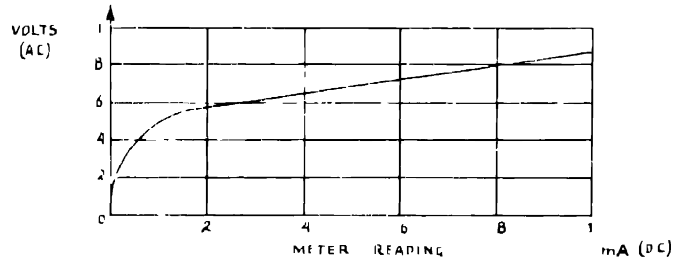


FIG. 247.—Voltage drop across AC meter and rectifier.

to the alternating current. It is most important, however, to note that, although no *current* is lost in the rectifier, a *voltage* drop is introduced. This voltage drop does not vary *linearly* with the current, so that the total impedance of rectifier plus meter is *not constant*. Fig. 247 shows the voltage drop across meter plus rectifier for a typical 1 mA rectifier meter with a coil resistance of 100 ohms.

This means that if the meter is used to measure the voltage across its terminals (*i.e.* is used as a voltmeter) the scale will not be linear at the bottom end.

Rectifier voltmeters

It has just been shown that, as the impedance of a rectifier meter varies with the current through it, it cannot be used as a voltmeter without recalibration. This is true only when small voltages have to be measured; when large voltages are to be measured, a large dropping resistance must be put in series with the meter, and this will "swamp" any variations in the impedance

of the meter, and hence give a linear scale. Thus, for example, with a 1 mA meter used to give a full-scale deflection (FSD) on 100 volts the drop across the rectifier plus meter is less than 1 volt, so the error cannot be greater than 1 volt on the scale.

Calculation of resistances for voltmeter

If the voltage drop across rectifier plus meter at full scale deflection is known, this can be subtracted from the required full-scale deflection voltage to give the required voltage drop in the series resistance. This resistance can then be calculated as follows :—

$$R = \frac{\text{Required voltage drop}}{1\cdot 11 \times \text{meter FSD current (DC)}}$$

(The denominator is of course the *alternating* current required for full-scale deflection.)

For example: consider the 1 mA meter already mentioned: a full scale deflection is required on 10 volts.

At full-scale deflection the meter voltage drop is, from Fig. 247, about 0·86 volt. Hence 9·14 volts must be dropped in the series resistor at full scale deflection, *i.e.* when the direct current is 1 mA and the alternating current is 1·11 mA (RMS).

$$\text{Hence} \qquad R = \frac{9\cdot 14 \times 1000}{1\cdot 11} = 8227 \text{ ohms}$$

The circuit is therefore as shown in Fig. 248.

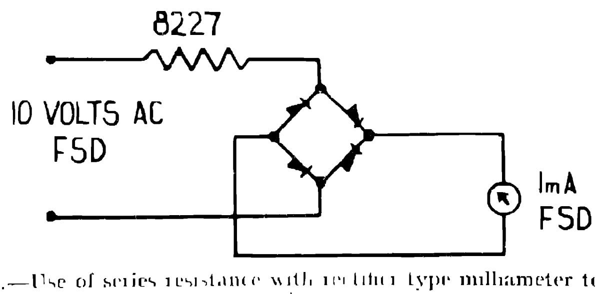


FIG. 248.—Use of series resistance with rectifier type milliameter to measure voltage

Note that there will be an appreciable error at low voltages, as 8227 ohms is not very large compared with the meter impedance. Thus when the meter reads 0·2, the alternating current flowing will be 0·222 mA, giving a drop of 1·83 volts across the 8227 ohms resistor. The voltage across the meter is (from Fig. 247) 0·57 volts, so the terminal voltage is 1·83 + 0·57 = 2·4, and not 2. In this case it would be necessary to provide another scale on the meter, or a calibration chart, for accurate measurements. Suppose, however, that the same meter was required to give full-scale deflection on 100 volts; the voltage drop in the meter at full-scale deflection is 0·9 volts, so $R = \frac{(100 - 0\cdot 9)}{1\cdot 11} \times 1000 = 89{,}190$ ohms.

In this case, when the meter reads $0 \cdot 2$ and the alternating current is $0 \cdot 222$, the voltage drop across R is $19 \cdot 8$, and that across the meter is $0 \cdot 57$, so the terminal voltage is $20 \cdot 4$, corresponding to an error of about 2 per cent. In this case the scale is reasonably accurate, and recalibration is probably unnecessary.

High current ranges

The range of a DC milliammeter is increased by the use of shunts; in AC meters, however, as the impedance varies with current, the multiplying factor of a shunt would also vary, and the meter would require separate scales for each range. For this reason, shunts are seldom found in AC meters; current transformers are used instead For example, suppose the 1 mA meter is required to read 100 mA, the turns ratio can be calculated as follows (*see* Fig. 249).

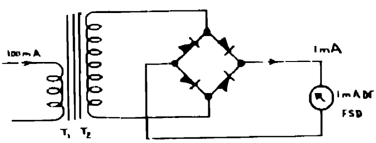


FIG. 249.—Use of current transformer to measure large currents.

The secondary alternating current required to produce 1 mA DC is $1 \cdot 11$ mA AC (RMS) and this is required when the primary current is 100 mA.

$$\text{Hence} \quad \frac{T_1}{T_2} = \frac{1 \cdot 11}{100} = 1 : 90.$$

Similarly, to read 1 amp, the ratio would be 1 : 900. The most important point in current transformer design is to keep the iron losses to a minimum. In the example given, the secondary would in practice probably be wound with 900 turns, giving 10 and 1 turns for the 100 mA and 1 amp primaries respectively. For measurements of larger currents, the primary often consists of a straight bar of metal with the secondary wound toroidally round it.

It is most important to ensure that the meter is never disconnected from the secondary while the primary current is flowing; for its removal might well increase the primary impedance 100 times, with a corresponding rise in secondary voltage. In many instances this is sufficient to destroy the transformer. If the meter were disconnected but the rectifier left in circuit, the rectifier would certainly be burned out.

Low voltage meters

It has been seen that rectifier meters do not give linear scales at low voltages. This can be overcome by using a transformer to step up the voltage and inserting a large resistance in series with the meter. Linearity is thus obtained, but at the expense of sensitivity.

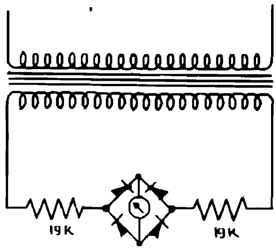


FIG. 250.—Use of voltage transformer.

An example of this is the decibel-meter (Fig. 250), which has to give a full-scale deflection on 0·775 volt. Here, a step-up ratio of about 1 : 2 is used, with a 40μA FSD meter and 38,000 ohms in series.

Frequency errors

For rectifier meters without transformers, these are negligible up to 100 kc/s. The performance of transformer meters depends on the design and construction of the transformer, but the response can usually be made flat over the audio-frequency range.

Temperature errors

These affect principally the voltage drop across the rectifiers, and can therefore be neglected in those circuits where this voltage drop is made to have little effect.

FURTHER APPLICATIONS OF RECTIFIERS

Biasing of rectifiers

It has been seen that the resistance of a rectifier depends on the DC voltage applied to it. When a DC voltage is applied in such

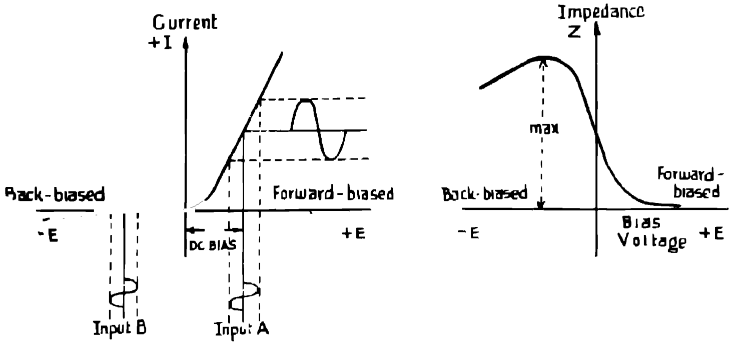


FIG. 251.—Variation of rectifier impedance with bias.

a direction that the rectifier offers a low resistance, the rectifier is said to be " forward-biased." If the voltage is reversed, the rectifier is said to be " back-biased."

Consider, in addition to a large forward-biasing voltage, the application of a small AC voltage (input *A*, Fig. 251). The current flowing will contain a large alternating component, and the rectifier therefore offers a low impedance to the AC voltage. The magnitude of this impedance depends on the slope of the current-voltage characteristic curve of the rectifier, and hence on the DC biasing voltage.

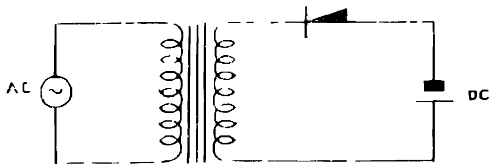


FIG. 252 Method of applying AC and DC independently to a rectifier.

If the DC biasing voltage is now changed, so that the rectifier is back-biased, and the same alternating voltage again applied (input *B*), very little current will flow, and the alternating component will be negligible. The rectifier therefore offers a high impedance to the AC voltage under these conditions.

Fig. 251 also shows the variation of impedance with bias voltage, it being noted that the rectifier offers its maximum impedance when an optimum bias voltage is applied.

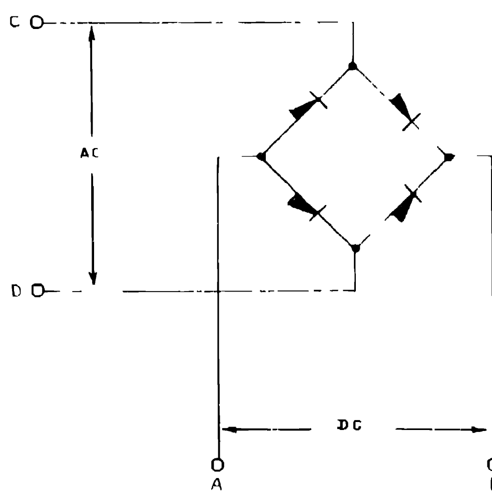


FIG. 253 —Alternative method of applying AC and DC independently to a rectifier network

One of the many ways of applying AC and DC separately to a rectifier is shown in Fig. 252. Another commonly found method is shown in Fig. 253; this circuit is in fact a full-wave rectifier bridge, but it is more convenient not to regard it as such. Clearly, if the DC bias is applied with *A* positive and *B* negative, the

rectifiers will all be forward-biased, and the AC impedance across *CD* will be low. If, however, *B* is positive and *A* is negative, the rectifiers will all be back-biased, and the AC impedance will be high.

These elementary circuits form the basis of many ingenious devices employed in line communication. Some common examples will now be considered.

Variable attenuators

A typical variable attenuator circuit is shown in Fig. 254. The input is applied across a " bridge " circuit having a rectifier W_1 in one arm, and a 250 $\mu\mu$F condenser in the other. The output from the other diagonal of the bridge appears across *A B*. If W_1 is back-biased, its impedance is high, and is balanced roughly by the 250 $\mu\mu$F condenser; in this case little output will appear across *A B*, and the attenuation will be high. If W_1 is forward-biased, the bridge will be unbalanced, and the attenuation low.

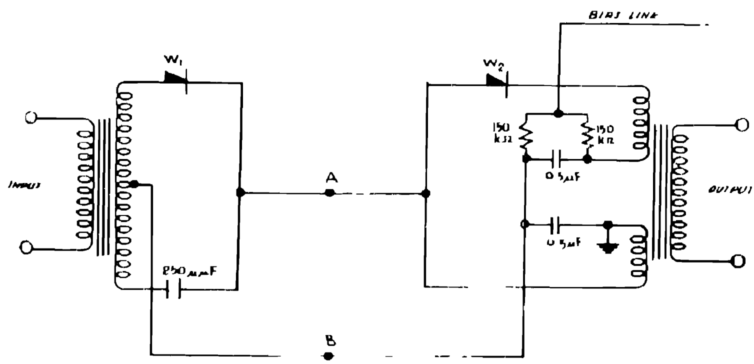


FIG. 254.—Typical variable attenuation circuit.

The output across *A*–*B* is applied to another bridge circuit consisting of a rectifier W_2 balanced in this instance by a short-circuit in the opposite arm. The other diagonal of this bridge is connected through a transformer to the attenuator output. This second bridge is roughly balanced when W_2 is forward-biased. Hence, for high overall attenuation, W_1 must be back-biased and W_2 forward-biased. For low attenuation, these biases must be reversed. By adjusting the bias to intermediate values, any desired attenuation may be obtained.

This is effected by applying a voltage to the " bias line ". If the DC paths are traced out, it will be seen that a positive potential to earth on the bias line causes W_1 to conduct and W_2 to be back-biased, giving low attenuation. On the other hand, a negative potential on the bias line reverses the biases, and gives high attenuation. The attenuation may thus be varied by adjusting the DC bias voltage.

Voltage limiters

Voltage limiters are designed to prevent the voltage at a point in a circuit from exceeding a certain peak value. This peak value is usually quite small, the commonest form of voltage limiter being the "crash limiter" used across telephone receivers. In its simplest form, this consists of two rectifiers back to back (*see*

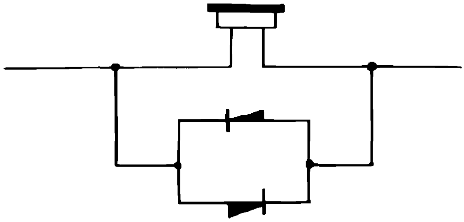


FIG. 255.—Use of rectifier as voltage limiter.

Fig. 255). When the voltage across the receiver exceeds 0·25 volt, the forward resistance of one or other rectifier drops, and shunts the receiver. This has the effect of limiting the voltage at that point.

It is possible to control the voltage at which limiting takes place by applying an initial DC back-bias to each rectifier, as in

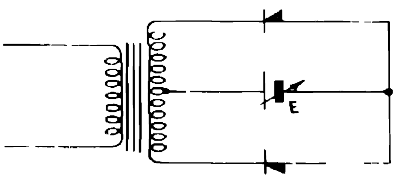


FIG. 256.—Biased voltage limiter.

Fig. 256. Here the voltage across each rectifier is E volts back-bias, plus the alternating voltage across half the transformer secondary. When the alternating voltage exceeds E, limiting will occur. By varying E, the level at which limiting occurs may be adjusted.

Certain materials have an impedance that drops as the applied voltage is increased in either direction, and these also are used as limiters. An example of this is the "ATMITE*" disc used in certain three-channel carrier telephone systems. Its current-voltage and impedance-voltage characteristics are shown in Fig. 257*a* and *b*.

Fig. 258 shows such a disc used as a voltage limiter at the input to the modulator on a multi-channel carrier telephone system. Its function is to prevent overloading of the transmitting equipment and consequent distortion; the effect on intelligibility is negligible. When the level of the input is 1 mW, the disc has high impedance and its shunting effect is nil. As the level increases, however, the resistance of the disc decreases and its shunt effect limits the level passing through the pad to the modulator.

* ATMITE is a Trade Mark owned by Automatic Telephone & Electric Co., Ltd., and is used by them to denote the particular non-linear resistance material which they market.

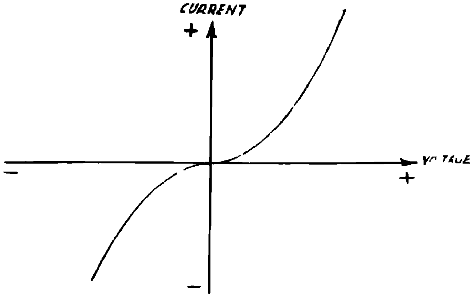


FIG 257*a* —Current voltage characteristic of an ATMITE disc

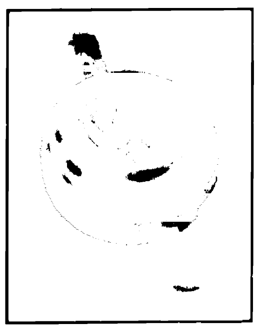


PLATE 14. ATMITE disc

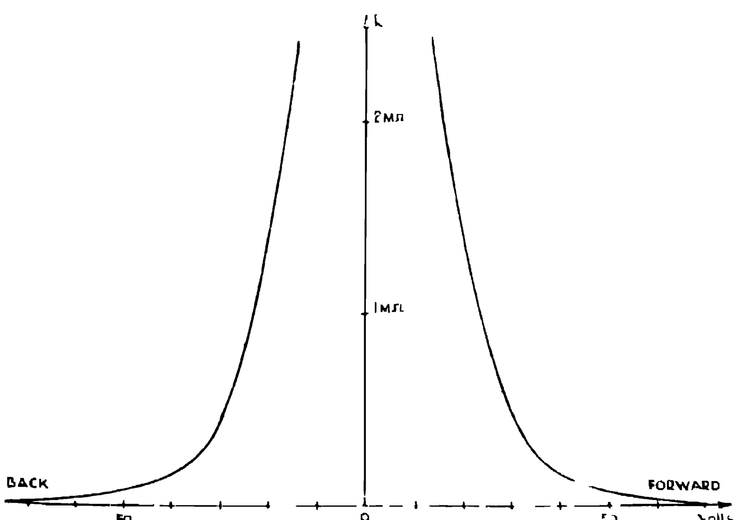


FIG 257*b* —Impedance voltage characteristic of an ATMITE disc

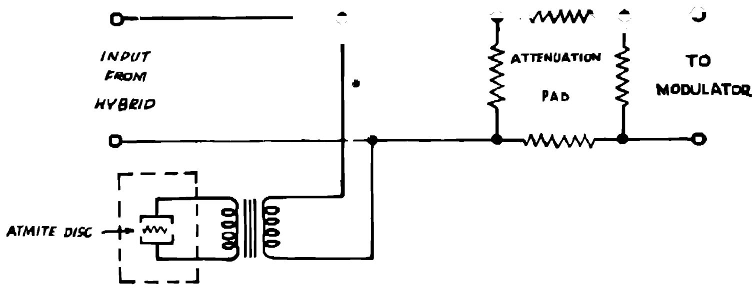


FIG. 258.—Use of ATMITE disc as voltage limiter.

The action of the limiter can best be seen from Fig. 259, which shows the voltage passing to the pad plotted against the voltage applied from a circuit having an impedance of 600 ohms with the limiter in and out of circuit.

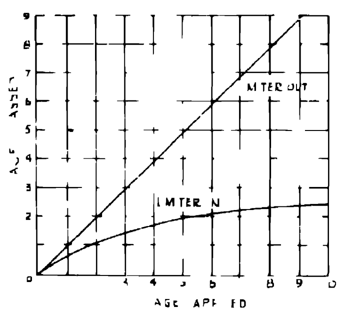
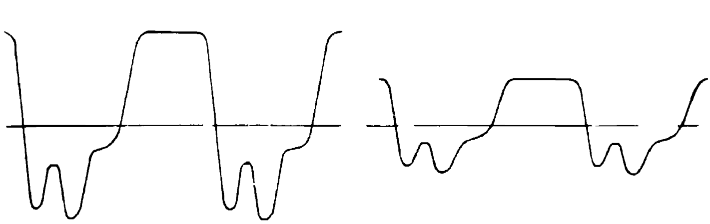


FIG. 259.—Action of ATMITE disc voltage limiter.

To show that the ATMITE disc causes little distortion, the output of a normal buzzer unit is given in Fig. 260. Fig. 260*a* shows the output of the buzzer, and Fig. 260*b* shows the same output after passing the limiter.



(a) BEFORE VOLTAGE LIMITER (b) AFTER VOLTAGE LIMITER

FIG. 260.—Effect of ATMITE disc on output of buzzer unit (obtained from CRO traces).

In addition to its use as a limiter, an ATMITE disc is occasionally used as a spark quench across relay contacts. By reducing voltage surges, it prevents sparking.

Relay slugging

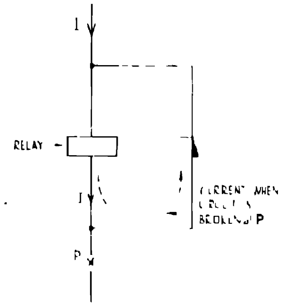


FIG. 261.—Rectifier as relay slug.

A rectifier in shunt with a relay coil can be used to make the relay slow-to-release. Consider Fig. 261; in the normal condition, with the relay operated, the rectifier is connected in such a direction that it will not shunt the relay. Without the rectifier, if the relay circuit be broken at P, the relay would release quickly. With the rectifier in place, however, the inductance of the relay tends to maintain a current through the low resistance of the rectifier. Until this current dies down, the relay will not release.

Meter shunts

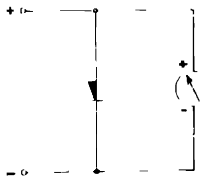


FIG. 262.—Rectifier as meter shunt.

Connected across a meter, a rectifier will act as a shunt whose resistance drops as the applied voltage increases. This has the effect of closing up the top end of the meter scale. With care, an approximately logarithmic scale can be obtained.

POWER SUPPLIES

Equipment in which thermionic valves are used always requires a DC high tension (HT) supply. This is not usually immediately available, but has to be derived from some other source—either AC mains or a low-voltage DC battery. For this purpose, some sort

of power supply unit has to be used. It is essential that the HT provided should have a minimum amount of AC present, and that the DC voltage should not vary appreciably when the current taken from the supply changes. The first problem (reducing the "ripple") is referred to as "smoothing"; the second (keeping the output voltage constant) is referred to as "regulation". These problems will now be discussed, together with various forms of power supplies.

Smoothing circuits

The output from a rectifier has been seen to consist of pulses of DC, and these can be shown, by Fourier's analysis, to consist of a steady DC component, equal to the mean value of the pulses, plus a large number of alternating components. For half-wave rectification, the DC component is equal to $\frac{1}{\pi}$ times the peak value of the pulses, and for full-wave rectification it is twice this value (*see* page 117) The alternating components, however, form a ripple that would be detrimental to the operation of most line equipment, and they are therefore removed by means of low-pass filters having a cut-off frequency lower than the lowest ripple frequency. These filters normally consist of one or more sections, each section comprising a choke or inductance in series with the rectifier output, and a condenser in parallel with the output, as shown in Fig. 263.

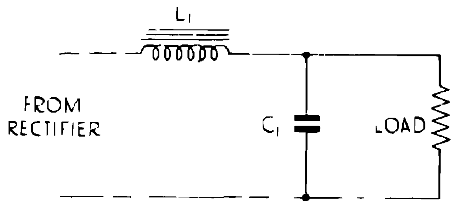


FIG 263 —Single section choke-input filter.

The theory of filters is dealt with in Chapter 15, but the operation of the smoothing filter now under discussion may be understood by considering the choke and condenser to form a potentiometer across the rectifier output At zero frequency (*i.e.*, for DC) the inductance offers a low impedance (merely that of its DC resistance), while the condenser offers an infinite impedance; the whole of the DC voltage developed by the rectifier is therefore applied to the load, and this voltage is equal to the mean value of the rectifier output. At the frequencies of the various ripple components, however, the inductance offers a high impedance, and the condenser a low impedance; only a fraction of the ripple voltage thus appears across the condenser, and therefore across the load. If particularly good smoothing is required, two or more such sections in tandem may be employed, as in Fig. 264.

If a "reservoir" condenser is connected in shunt across the rectifier output before the first choke, as C_1 in Figs. 266 and 268, the resulting smoothing network is called a "condenser-input" filter. On no load, a condenser-input filter gives a larger output voltage

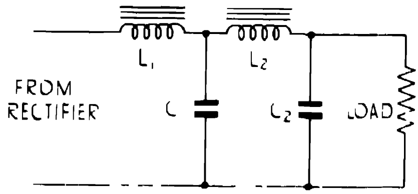


FIG. 264.—Double section choke-input filter.

than a choke-input filter. This is because, on no load, the condenser charges up to the peak (not the mean) voltage of the pulses; and, retaining this voltage from the peak of one pulse to the next, gives a DC output voltage that is equal to the *peak* value of the alternating voltage applied to the rectifier. On load, the condenser partially discharges through the load during the periods between pulses (*see* Fig. 265), and the mean DC output voltage drops. Thus the condenser-input filter gives a higher output voltage on light loads than the choke-input filter, but the "regulation" is not so good; that is to say, the DC output voltage drops appreciably with increase

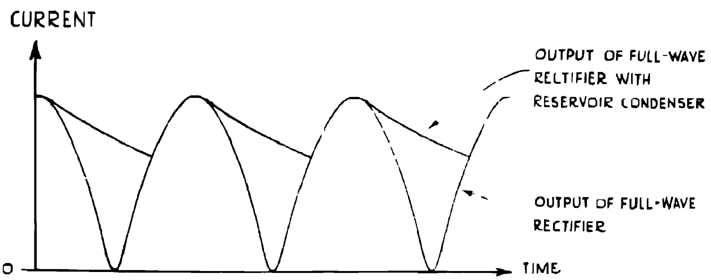


FIG. 265.—Output of full wave rectifier with reservoir condenser

in load. The condenser-input filter has the advantage that the reservoir condenser itself assists in the smoothing, and hence a reservoir condenser followed by *one* choke-and-condenser section will give a degree of smoothing comparable with that obtained from *two* choke-input sections.

The values of the components for filters of either the choke-input or the condenser-input type can be calculated by the following empirical methods.

Design of condenser-input filters

(1) *For half-wave rectifiers*

(*a*) Knowing V (the RMS voltage across secondary of transformer), and V_0 (the required direct voltage across the

load), calculate $\frac{V_0}{V}$. Let the required direct current through the load be I_0. Calculate $R\left(=\frac{V_0}{I_0}\right)$

(*b*) Using graph *a* of Fig. 267, read off $\omega C_1 R$ ($\omega = 2\pi \times$ supply frequency), and thus obtain the value of C_1.

(*c*) From graph *b* of Fig. 267 read off the percentage ripple $\frac{V_R}{V_0}$ across C_1. (V_R is the RMS ripple voltage.)

(*d*) Calculate L_2 and C_2 from the formula:—

$$\frac{\text{Per cent ripple across } C_2}{\text{Per cent ripple across } C_1} = \frac{1}{\omega^2 L_2 C_2}$$

the aim being to reduce the percentage ripple across C_2 to a minimum (e.g. 0.2 to 0.3 per cent if possible).

Units used are ohms, farads, and henries.

(2) *For full-wave rectifiers*

(*a*) Knowing V, V_0 and I_0 calculate $\frac{V_0}{V}$ and R.

(*b*) Using graph *a* of Fig. 269 read off $\omega C_1 R$ ($\omega = 2\pi \times$ supply frequency) and thus obtain the value of C_1.

(*c*) From graph *b* of Fig. 269, read off the percentage ripple across C_1.

(*d*) Calculate L_2 and C_2 from the following formula:—

$$\frac{\text{Per cent ripple across } C_2}{\text{Per cent ripple across } C_1} = \frac{1}{4\omega^2 L_2 C_2}$$

the aim being to reduce the ripple across C_2 to a minimum.

In step (*d*) for the full-wave rectifier the $4\omega^2$ is introduced because the ripple frequency is now twice that of the supply.

Design of choke-input filters

Consider the single-section filter shown in Fig. 263. Using a full-wave rectifier circuit, the values of L_1 and C_1 can be obtained from the formula

$$\text{Percentage ripple at output} \simeq \frac{144}{L_1 C_1}$$

This is an approximate formula but the values obtained will give suitable smoothing for 100 c/s ripple. If the ripple frequency is not 100 c/s then values obtained for L_1 and C_1 should be multiplied by $\frac{100}{f}$ where f is the actual ripple frequency.

Condenser values so obtained are in microfarads, and inductance values in henries.

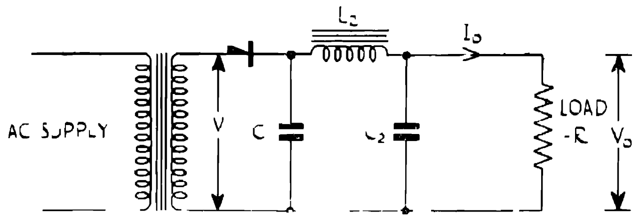


Fig 266 —Half-wave rectifier with condenser input filter

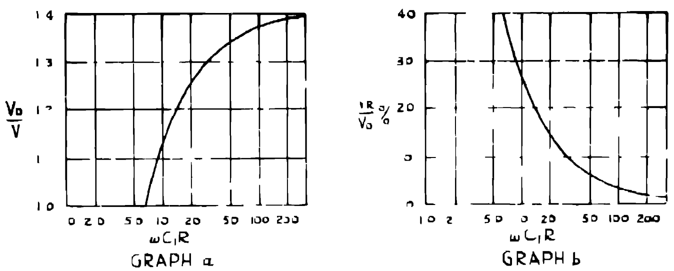


Fig 267 Design data for a condenser input filter for use with a half wave rectifier

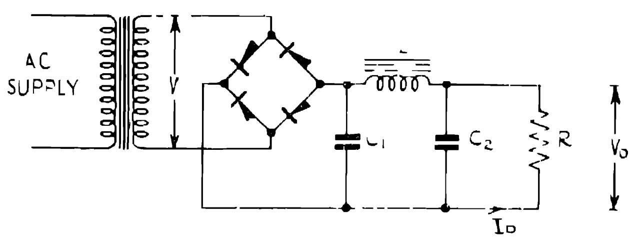


Fig 268 — Full wave (bridge) rectifier with condenser input filter

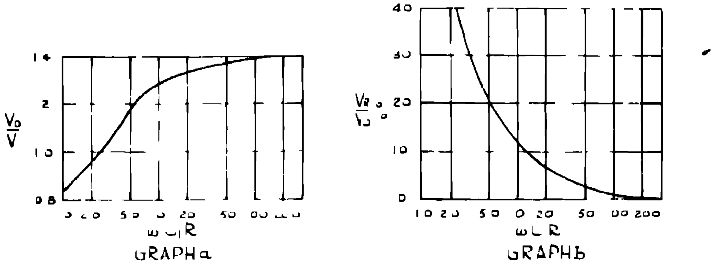


Fig 269 — Design data for a condenser input filter for use with a full wave rectifier, assuming a fundamental ripple frequency of 100 c/s

The distribution of inductance and capacity (only the product L_1C_1 has been obtained) depends on the size of choke to be fitted. There is a critical value to L_1, however, and a choke of lower value will lack the good voltage regulation which is the advantage of filters of this type

$$L \quad \text{(in henries)} \simeq \frac{\text{load resistance (in ohms)}}{1000}$$

To obtain a small percentage ripple it is better to use a double section filter, as component sizes would be more economical.

The values of L_1, L_2, C_1 and C_2 can be obtained from the formula :

$$\text{Percentage ripple in output} \simeq \frac{1350}{L_1L_2(C_1+C_2)^2}$$

This again gives merely a guiding relationship between the various components and actual values will be chosen to prevent too great a voltage drop and to give suitable and economical components The units are capacity in microfarads and inductance in henries

Resonance—If series resonance occurs in L_1 and C_1 of the filter, large alternating voltages will build up and the reverse of smoothing will result To avoid this, the product of the inductance in henries times the capacity in microfarads should not be lower than 5 in the case of a 50 c/s supply For this supply frequency, resonance occurs when L_1C_1 2 53, but a large safety margin is essential

POWER SUPPLIES WORKING FROM DC

The preceding paragraphs have shown how a power supply can be made to operate from AC mains If, however, the source of power is a low-voltage DC battery some means must be found of stepping this up to the required HT voltage The two most common ways of doing this are by using either a vibrator or a rotary transformer.

Vibrators

A vibrator is used to change the DC to a low-voltage AC supply, which can be stepped up by a transformer and dealt with as before. The way in which a vibrator does this is described below

The shunt-drive type vibrator

Fig 270*a* shows the simplified circuit of a shunt-drive type vibrator In the "idle" position, the armature does not make with either contact but when the DC supply is connected, magnetising current flows through the coil attracting the armature to the top contact This results in DC flowing through the top half of the transformer primary winding At the same time this contact shorts out the operating coil and after the magnetic flux has decayed the armature is released, the momentum of the latter carries it past its central position and on to the bottom contact. DC now passes through the lower half of the transformer in the

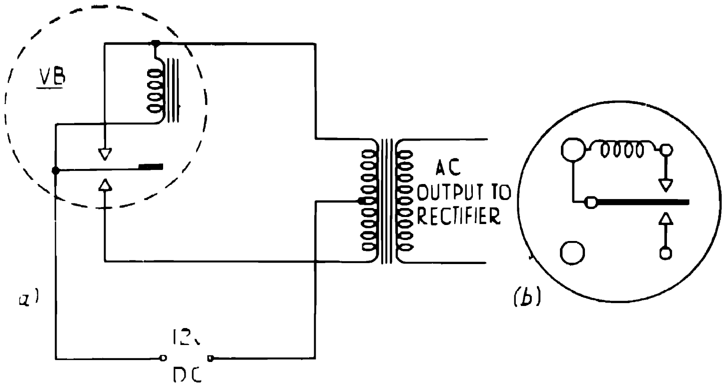


FIG. 270 —Shunt drive type vibrator

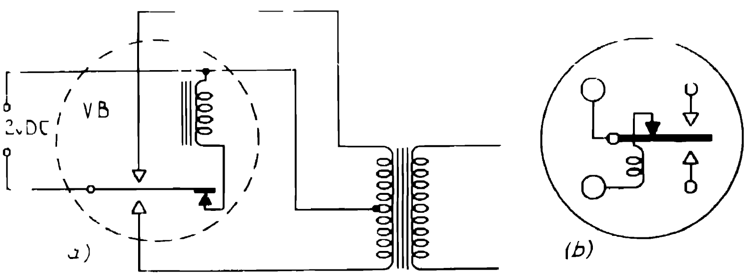


FIG. 271 —Series drive type vibrator.

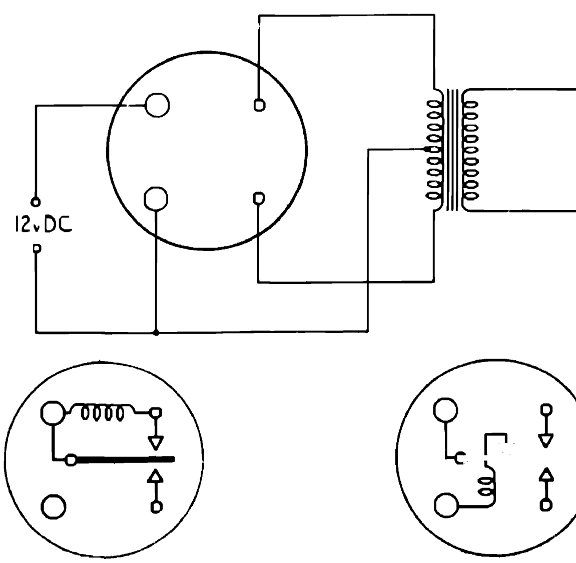


FIG 272 —Circuit for shunt or series drive vibrator.

reverse direction to the previous pulse. Thus, there occurs, in the transformer, a reversal comparable with the two halves of a cycle of AC. Meanwhile the coil is again magnetised, and the armature keeps vibrating between the contacts until the DC supply is disconnected. The constant reversals in the primary induce an alternating EMF into the secondary; if a suitable turns ratio be chosen, the required HT voltage is obtained, and can be rectified as described in previous sections. Fig. 270*b* shows the diagrammatic representation of the shunt-drive type vibrator.

The series-drive type vibrator

In this later type of vibrator, the coil is in series with a third contact that is making with the armature in the rest position (*see* Fig. 271*a*). When the supply is connected, current flows through the coil and causes attraction of the armature. The armature then makes contact with the upper contact and DC flows through the top half of the transformer primary. This movement also breaks the circuit of the coil, so that the armature is released; and it therefore travels back through its rest position and, owing to its momentum, continues on to the lower contact. The armature is now in contact with both the lower interrupter contact and the coil contact. Thus a DC pulse passes through the lower half of the transformer and the coil is energised once more. The armature returns to the top contact and the cycle begins again.

The current in the primary thus consists of reversals of DC which, as before, have the same effect as ordinary AC and produce the required high alternating voltage across the secondary.

Fig 271*b* shows the diagrammatic representation of the series-drive type vibrator.

Vibrator circuit arrangements

In the case of the shunt-drive type of vibrator, although four pins are fitted on the base, only three are used, whereas in the series-drive type all four pins are employed; circuits may, however, be so arranged that either type may be inserted (*see* Fig. 272).

When the vibrator contacts make and break the DC circuit, heavy induced voltages are set up which may cause sparking at and damage to these contacts. To remove this possibility spark quench condensers are fitted between the armature and the two contacts, as shown in Fig. 273*a*. A typical value for these condensers is 0·01 μF. Sometimes a "buffer" condenser is fitted across the secondary of the transformer, as shown in Fig. 273*b*. The action of this buffer condenser is principally to improve the commutation, thereby increasing the efficiency and prolonging the contact life, but it also performs the function of the spark quench condensers, as it offers an easy path for the heavy surges.

Synchronous vibrators

The synchronous, or self-rectifying, vibrator circuit is arranged to reverse the polarity of the secondary at the same instants as

(a)

(b)

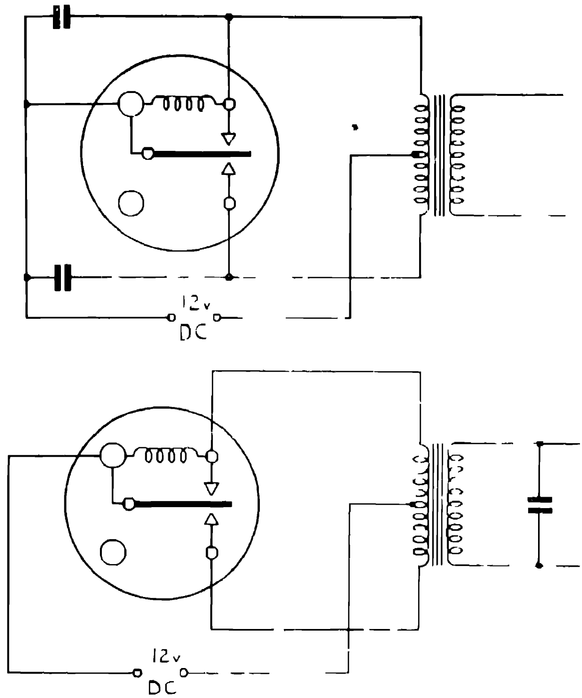


FIG. 273.—Vibrator circuit showing (a) spark quench condensers, (b) buffer condenser.

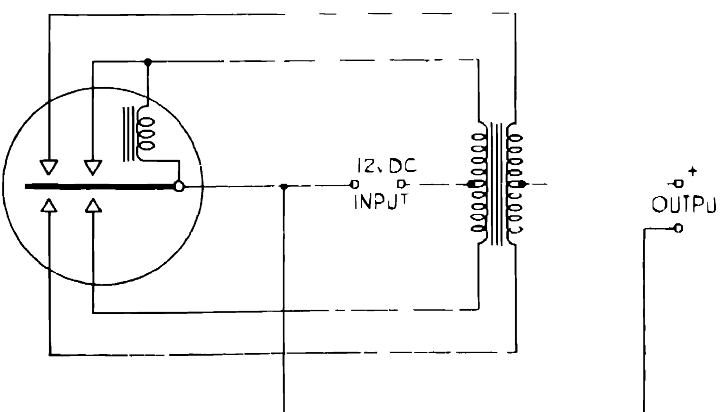


FIG. 274.—Synchronous vibrator.

the reversals occur in the vibrator. The output is therefore rectified and consists of pulsating DC which can be smoothed by the circuits already covered.

The main difference from the non-rectifying vibrator circuit is the provision on the vibrator of additional contacts, which are connected to the terminals of the secondary coils (*see* Fig. 274). When the armature makes with the lower contacts DC builds up through the lower half of transformer primary, inducing an EMF in the lower half of the secondary. By controlling the polarity of the LT supply, it can be arranged that the centre tap of the secondary assumes a positive potential with respect to the lower terminal which is now earthed through the vibrator. The next induced EMF, due to the break of this DC and the make of the DC in the other half of primary, will be reversed; but, owing to the change-over of the secondary contacts, it will be taken from the top half of the secondary, so that the centre tap remains of positive polarity.

Thus by synchronising the change-over contacts, pulsating DC is obtained in the output, the centre tap of the secondary winding being the positive terminal. This type of vibrator assembly, by avoiding the need for rectifiers, is economical, eliminates the loss of energy in the rectifiers, and eliminates the rectifiers as a source of faults. This last advantage is, however, off-set by the fact that synchronous vibrators are more liable to faults than non-synchronous, and the overall fault-liability of both systems is about equal.

The contacts of the synchronous vibrator are slightly staggered so that the secondary contacts break a little before the primary and make a little after. This removes the danger of the secondary surge, due to the primary break, being passed to the output with the wrong polarity.

Radio interference suppression

The transient voltage surges in vibrator circuits contain radio frequency noise components, which cause interference in any near-by radio equipment. This noise cannot be controlled by the spark quench or buffer condensers, and other means of avoiding it must therefore be incorporated. The methods of suppression used in the past have not been standardised, and the circuits used are those which were found most suitable in the particular cases. The methods adopted incorporate :—

(1) Shielding—both magnetic and electrostatic;

(2) Good earthing;

(3) RF filtering in the leads to and from the vibrators.

All these combined give fair, but not full, suppression of the interference.

Fig. 275 shows an example of such a unit, typical values being :—

C_1	..	0·1 μF	C_2	..	0·25 μF
L_1	..	75·0 mH	L_2	..	32 μH

Radio frequencies set up at the contact of the vibrator driving

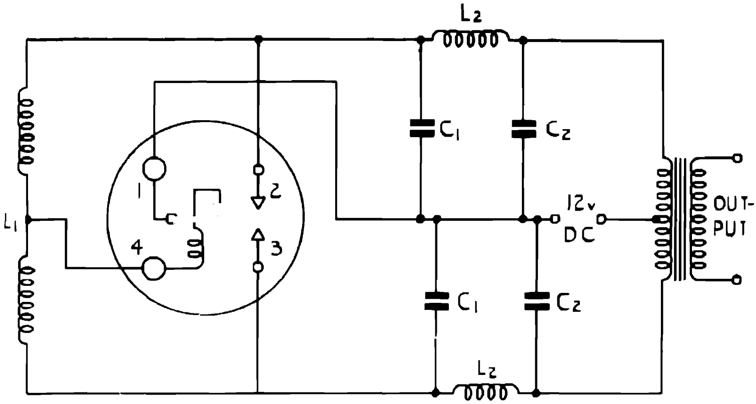


FIG. 275.—Radio interference suppression.

coil are suppressed by L_1, while those from the main contacts are suppressed by the low-pass filter formed by L_2, C_1, and C_2, which has a cut-off frequency well below 1 Mc/s.

Equally as important as the filter is the thorough shielding of the power supply and its connecting leads, since even a small piece of wire or metal will radiate sufficiently to cause interference in a sensitive receiver.

Mechanical construction of vibrators

The metal can of the vibrator supplies the requisite screening, while in the later models it also provides hermetic sealing. Without

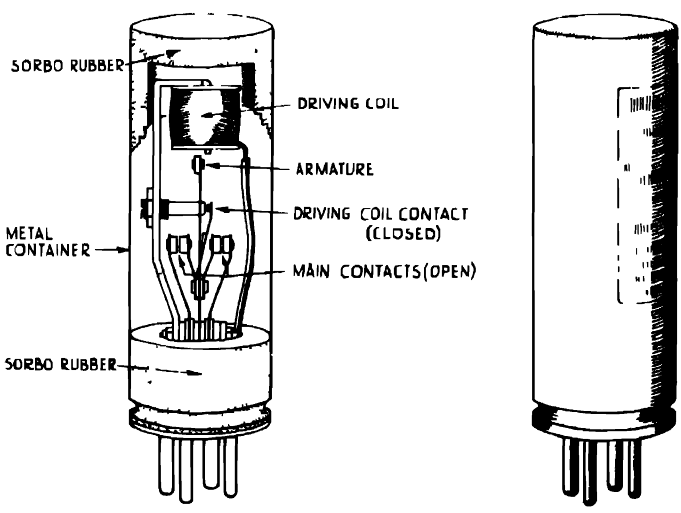


FIG. 276.—Construction of series-drive type vibrator.

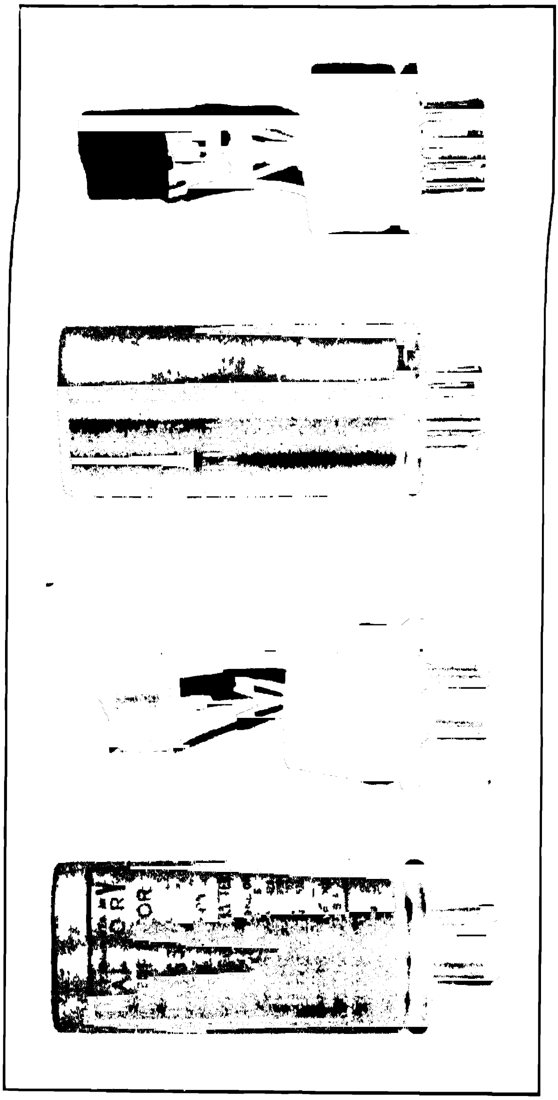


PLATE 15.—Vibrators.

this latter the vibrator may easily stop in a humid or rarefied atmosphere. In the tropics, damp entering the can forms a film on the contacts which prevents metallic contact ; thus no current can flow and the vibrator stops. In rarefied atmosphere the vibrator contacts start sparking at low voltages, so that the contacts deteriorate. Both of these difficulties are removed by hermetic sealing. When stoppages occur due to humidity a cure, though drastic, can be effected by force-driving the vibrator off 230 volts AC mains in series with a 30 watt lamp. This voltage breaks down the film, and the contacts are cleared for normal operation.

The frequency of the vibrator reed is not standardised, but, in the types used by the army, is of the order of *100 c/s*. It is governed principally by the mechanical inertia of the reed and the electrical characteristics of the coil ; and thus, in the miniature vibrators now being developed, a higher frequency is used.

Rotary transformers

Rotary transformers consist in effect of a DC motor and generator, with their armatures coupled mechanically. The motor works from a low-voltage supply, usually 12 volts or 24 volts, and the generator is designed to produce the required HT voltage.

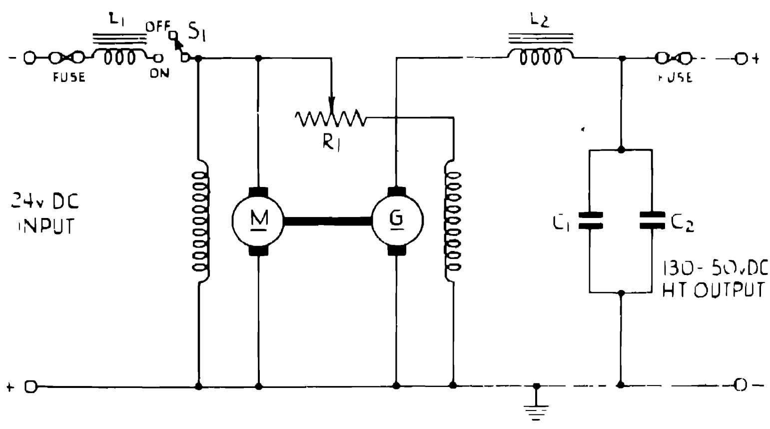


FIG. 277.—Circuit of rotary transformer.

In some cases, the two armatures are wound together and work in the same field. In other cases, the two fields are quite separate ; an example of this type will now be considered.

The output of the supply is :—

HT 130–150 volts, maximum drain 285 milliamps.

It is designed to operate from 24 volts DC.

Fig. 277 shows the circuit of the rotary transformer. L_1 is a choke fitted to prevent AC caused by commutator ripple from

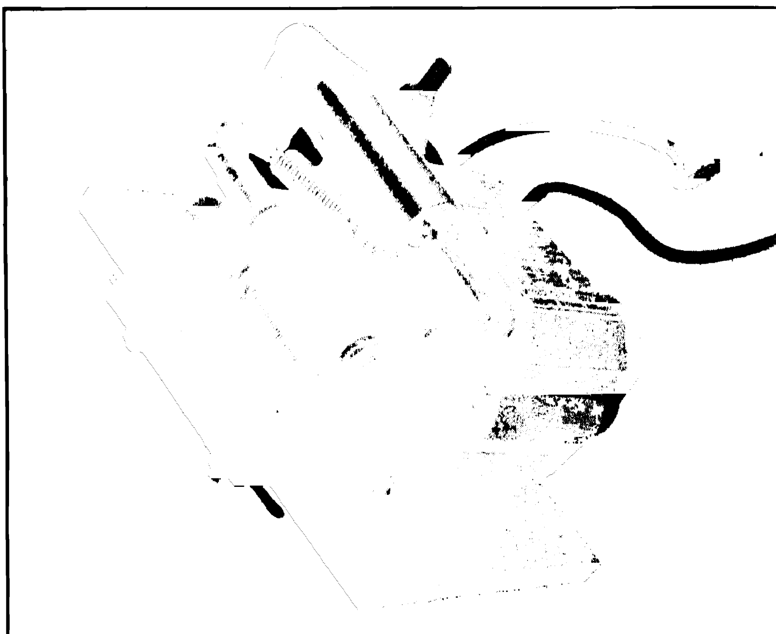


PLATE 16.—Carbon pile regulator.

feeding back to the supply and thus rendering it noisy. L_2, C_1, and C_2 form the smoothing circuit. The output voltage is controlled by R_1, which varies the current in the field coil of the generator.

If required, a carbon pile regulator may be fitted to ensure a constant HT voltage output for all loads up to the maximum.

Fig. 278 shows the circuit of such a regulator, and Fig. 279 a diagrammatic arrangement of its mechanical construction.

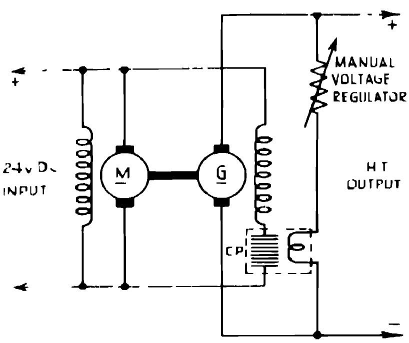


FIG. 278.—Regulation by use of carbon pile.

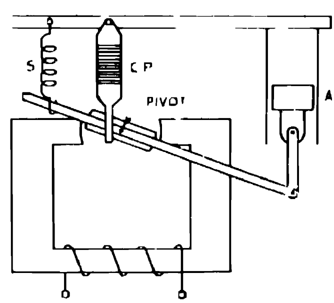


FIG. 279.—Carbon pile regulator.

The carbon pile consists of a stack of carbon plates, the resistance of which is varied by the pressure exerted on it. Normally it is held under pressure by the spring S; but when HT is being supplied, current flows through the coil of the electro-magnet, and the field between the pole-pieces tends to rotate the soft iron armature against the tension of the spring, and increases the resistance of the carbon pile.

In operation a state of stability is reached, and the carbon pile will be under a certain pressure due to the normal value of output voltage. When the load current increases, the HT voltage tends to fall, due to the increased armature voltage drop. This drop in voltage output causes less current through the electro-magnet of the carbon pile regulator and therefore decreased resistance. The resultant increase in generator field current produces the required increase in generator EMF.

POWER SUPPLY UNITS

The main parts of any power supply have now been considered; to illustrate their use, several complete power supply units will be discussed.

Example 1

A power supply unit providing 80 + 80 volts and 12 volts DC, from either 12-volt DC or AC mains.

The design (Fig. 280) is conventional; the AC is rectified by a voltage doubler circuit, and the output smoothed by reservoir condensers. The LT supply is rectified by a full-wave bridge, and again smoothed by a reservoir condenser. A high degree of smoothing is not required for either supply.

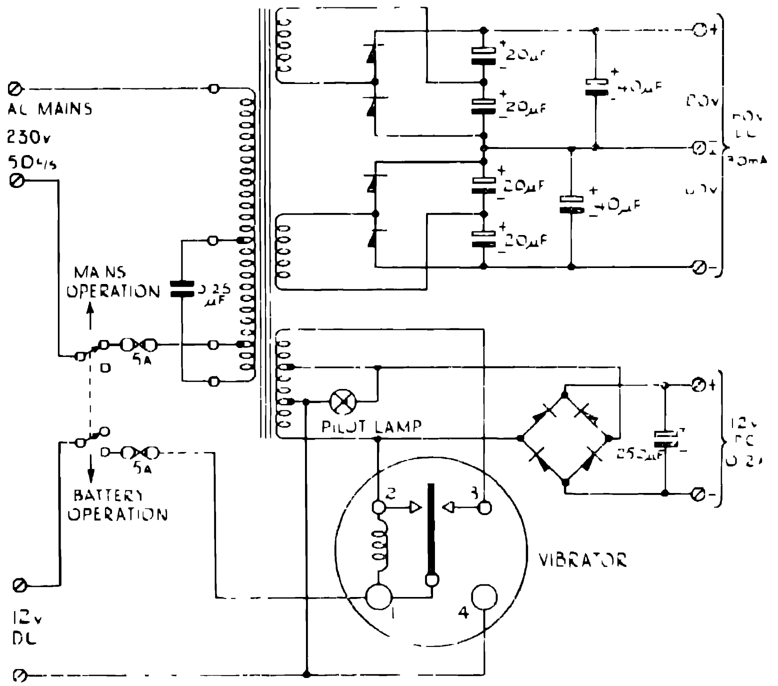


Fig. 280.—Power supply circuit providing 80 + 80 V, and 12 V DC from AC mains or from 12 V DC.

When working off a 12-volt battery, the LT terminals will not be used, since 12-volt DC may be obtained directly from the battery. The 0·25 μF condenser forms a buffer condenser for the vibrator.

Example 2

A power supply unit providing 130 volts HT and 24 volts LT from AC mains, with arrangements for a standby supply.

This unit (*see* Fig. 281) is designed to work off various supply voltages, the mains being connected to the appropriate terminals on the terminal strip shown on the left of the diagram. The neon lamp lights when the mains supply is in use. The remainder of the circuit is standard, apart from the relay *A* and the AC contactor switch *CS*. These cause automatic change-over to the standby supply, should the mains supply fail.

Relay *A* is across the HT supply and is normally operated. This causes operation of *CS*, placed across the LT supply, and this contactor completes the circuit for normal working. Should the LT fail, *CS* releases; should the HT fail, *A* releases, again causing *CS*

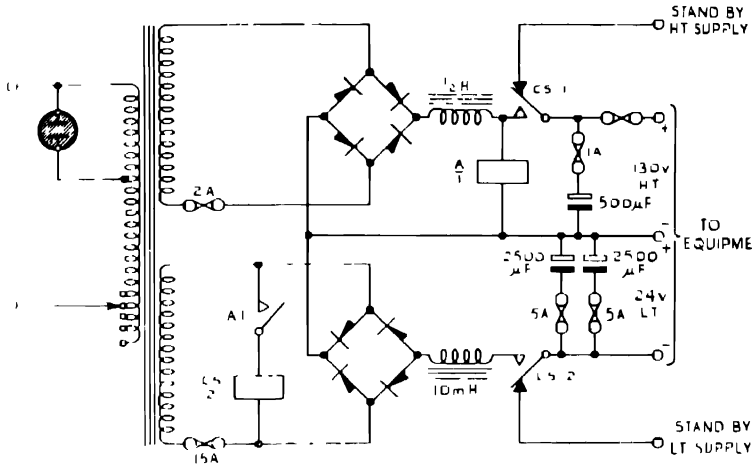


FIG. 281.—Power supply unit providing 130V HT and 24V LT from AC mains, with provision for standby supply.

to release. Thus, if either supply fails, the contacts of *CS* will release, thereby connecting the standby supply to the output terminals.

The outputs of this supply unit are :—

LT : 24 volts, 3 to 11 amps.

HT : 130 volts, 150–600 milliamps.

Example 3

A power supply unit providing 12 + 12 volts DC, at 6 amps, from AC mains.

This unit (*see* Fig. 282) is designed for use off 50 c/s AC mains, the necessary stepping up from the lower voltage supplies being achieved by an auto-transformer. The two subsequent pairs of transformers each change the single-phase input to a three-phase output which is designed to give a steadier rectified voltage than that possible with rectified single-phase voltages. The output is then fed to two sets of terminals; the first through a smoothing circuit to give an output suitable for telegraph supply, and the second without any smoothing giving a supply suitable for driving a motor. This double source prevents interference from the motor feeding back to the telegraph circuit. The output voltage in each case is 12 + 12.

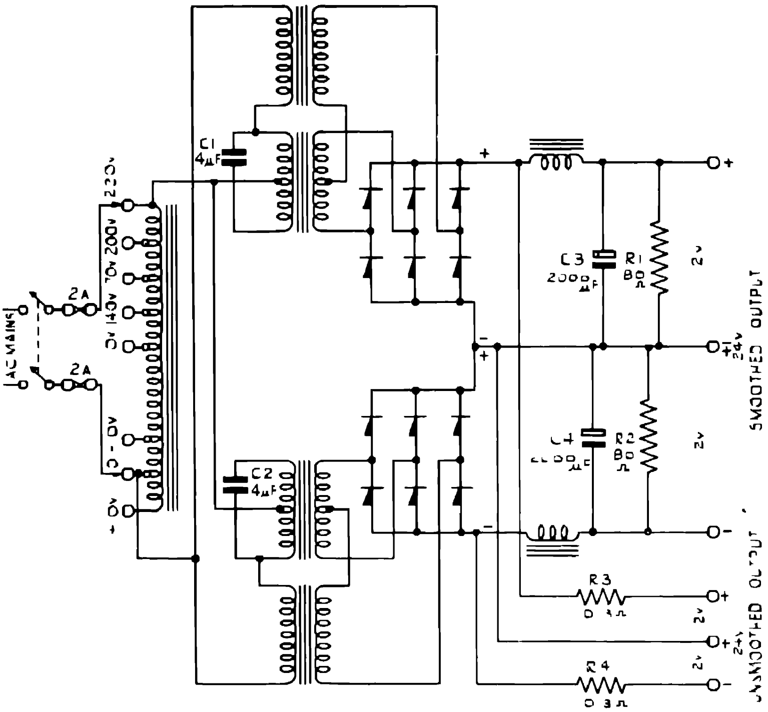


FIG 282—Power supply unit providing 12 V + 12 V DC at 6 amps, from AC mains

To understand the conversion to three-phase, consider the circuit of Fig. 283. The transformer T_1 has an inductive primary impedance that is largely resistive on full load On no-load, its impedance is largely inductive, so that I will lag behind V_1 by

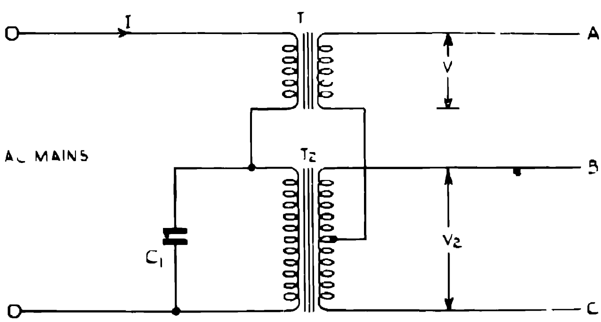


FIG. 283.—Method of obtaining three-phase supply.

almost 90°. Taking the supply current as reference vector, Fig. 284 shows the voltage across T_1 on full load; the arrow shows how the voltage vector rotates as the load is reduced.

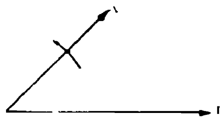


FIG. 284.—Vector diagram for voltage across primary of T_1 on full load.

Consider now the lower transformer; if C_1 is sufficiently large, the parallel combination of T_2 and C_1 will have a capacitive impedance, so that V_2 will lag behind the mains current I, as shown in Fig. 285. By a suitable adjustment of C_1, V_2 can be made to lag behind V_1 by exactly 90° on full load.

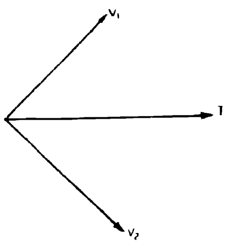


FIG. 285.—Vector diagram for voltage across primary of T_2 on full load.

Consider now the voltages of the three terminals A, B, and C relative to the mid-point O of the transformer T_2. The voltages V_1 and V_2 are redrawn in Fig. 286, with V_1 as reference vector, and this is the voltage of A (relative to O). The voltage of B is in phase with

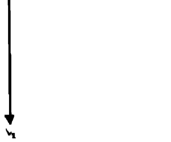


FIG. 286.—Vector relationship between voltages across the two transformer primaries on full load.

V_2, but equal in magnitude to $\frac{1}{2}V_2$; and the voltage of C is 180° out of phase with V_2 and equal in magnitude to $\frac{1}{2}V_2$ (*see* Fig. 287*a*). The voltages between the three terminals A, B, and C (which are shown in their cyclic order in Fig. 287*b*), are therefore given by :—

$$V_{AB} = V_{AO} + V_{OB} = -V_{OA} + V_{OB} = -V_1 + \tfrac{1}{2}V_2$$
$$V_{BC} = V_{BO} + V_{OC} = -V_{OB} + V_{OC} = -\tfrac{1}{2}V_2 - \tfrac{1}{2}V_2 = -V_2$$
$$V_{CA} = V_{CO} + V_{OA} = -V_{OC} + V_{OA} = +\tfrac{1}{2}V_2 + V_1 = V_1 + \tfrac{1}{2}V_2$$

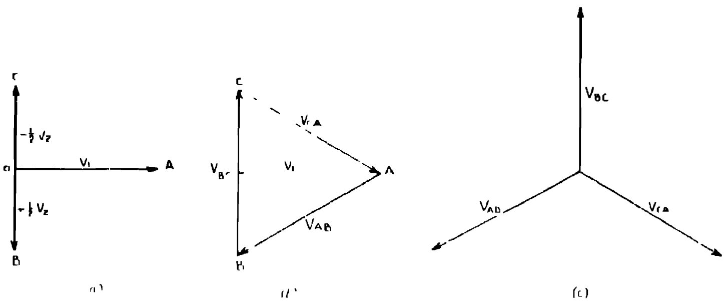


FIG. 287.—Complete voltage vector diagram for three-phase conversion.

It can be seen that if $\frac{\frac{1}{2}V_2}{V_1} = \frac{1}{\sqrt{3}}$, *i.e.* if $V_1 = \frac{\sqrt{3}}{2} . V_2$, then the triangle will be equilateral; this relationship can be satisfied by suitable choice of transformer turns ratios. The three points A, B and C then form a three-phase supply, the vector voltages between AB, BC, and CA being equal in magnitude. If the vectors representing these voltages are drawn in "star" form, as in Fig. 287*c*, it can be more clearly seen that they are separated by angles of 120°.

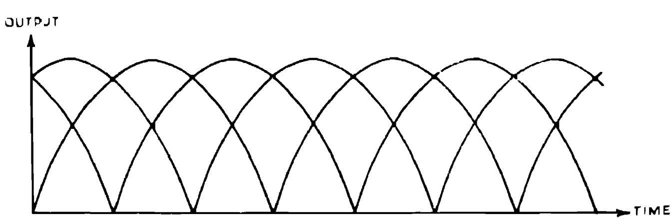


FIG. 288.—Output from three-phase full-wave rectifier.

This three-phase supply is applied to a full-wave bridge rectifier, the output voltage of which is equivalent to the combined effect of three single-phase rectified outputs differing by 120°. This gives a wave-form as shown in Fig. 288.

It can be seen that the lowest frequency present is six times the supply frequency, and hence smoothing is a simpler problem than with a single-phase supply.

This, however, is not the most important advantage of this system; it has also a constant output voltage at all loads. In a normal rectified power supply, the output voltage drops as the load is increased, due to resistance losses in transformers and smoothing circuits, and increased drops in the rectifier. In this case, however, the circuit is designed to give efficient conversion to three-phase at full load. As the load drops, the voltages V_1 and V_2 vary in magnitude and phase, which causes the rectified voltage to drop. If this drop as the load is reduced is made to balance the decreased resistive drop, a constant output voltage can be obtained for wide variations of load. This does not apply at zero load, so to prevent any trouble arising from this, a constant initial load is provided by the two 80-ohm resistances across the output. The net result is a constant output voltage for all loads up to 6 amps.

CURRENT AND VOLTAGE STABILISERS

In many circuits, it is necessary that the voltages and currents applied to the equipment be kept within close limits, while the supply may fluctuate over a wide range. For this reason, many stabilising devices have been developed.

Current stabilisers—the barretter

Current stabilisers are designed so that an increase in the current through them causes an increase of their resistance. If therefore they are placed in series with the load, their effect will be to stabilise the value of the current.

The most common form is the barretter, which is a lamp with a special filament—usually pure iron drawn to a fine wire—in a bulb

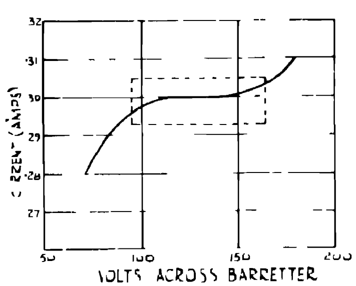


FIG. 289.—Typical voltage-current characteristic of barretter.

filled with hydrogen. The reasons for gas filling are to prevent oxidation of the filaments and to permit rapid removal of heat and therefore speedy stabilisation.

Fig. 289 shows the voltage-current characteristic of a typical

barretter working at 0·3 amps over a voltage range of 95 to 165 volts. This voltage range is called the working or " barretting " range of the barretter, and varies with the different designs. The working current level can be altered by design between limits of 0·2 amp and several amps. The former limit is governed by the possible fineness of the wire filament, while the latter is governed by the physical size of the lamp, a large size being necessary to dissipate the heat produced. This heat will be wastage of power and, as this is never desirable, the use of a barretter should always be carefully considered before adoption.

In operation, the barretter should be allowed ample air-circulation, as the ambient temperature naturally affects its working. It should also be removed or shielded from any strong magnetic field, as this would induce noise into the circuit.

Voltage stabilisers

In contrast to the current stabiliser, whose impedance increases with current, voltage stabilisers have an impedance which decreases with increase of voltage. For this reason they are used in parallel with the load. When the supply voltage increases, the impedance of the stabiliser will drop owing to the increased voltage across it.

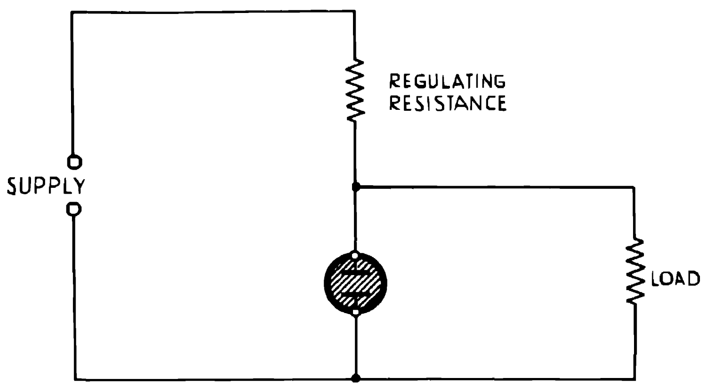
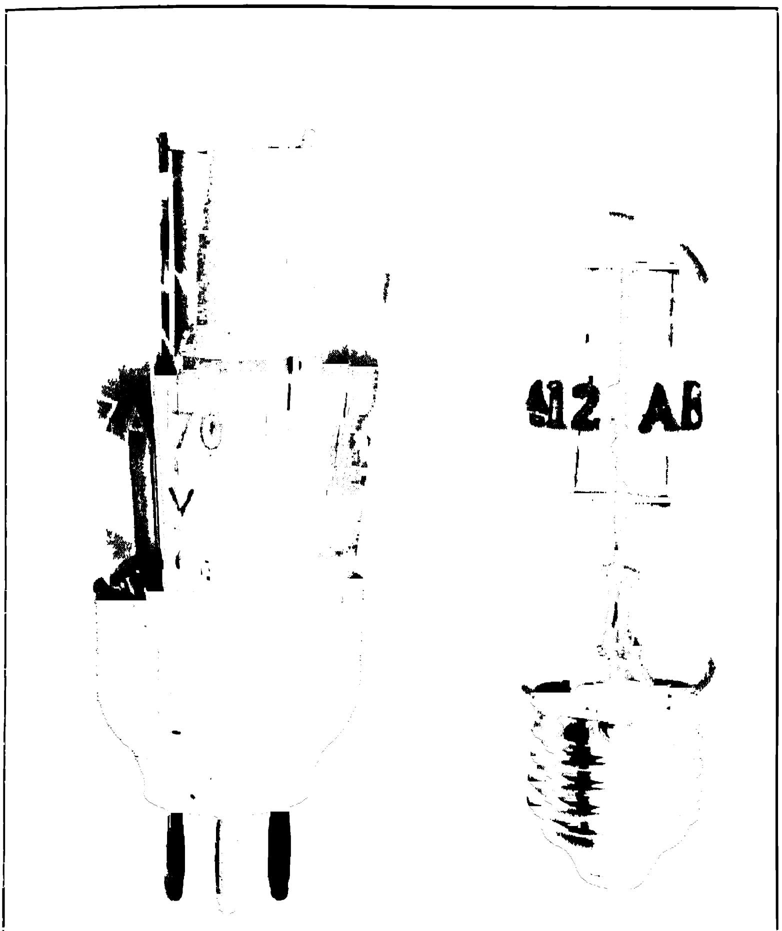


FIG. 290.—Voltage stabiliser circuit.

This causes a greater proportion of the current to flow through the stabiliser and less through the load. A circle of changes will therefore be set up which will tend to keep a constant voltage across the load, and therefore a constant current through it.

A common type of voltage stabiliser is the cold-cathode neon tube. The epithet " cold-cathode " is given because the cathode requires no heater current. The voltage builds up across the electrodes until it is high enough to cause the tube to " strike ". This striking voltage may be anything from 80 to 180 volts, and is usually about 100 volts in the case of the tubes used in line equipment.



(a) (b)

PLATE 17.—(a) Neon stabiliser
(b) Barretter

Once the tube has " struck ", its resistance falls abruptly from infinity to a very low value, so that the direct application of the striking voltage would cause a very large discharge current—large enough, in fact, to cause destruction of the tube. To avoid this, a resistance is connected in series with the tube as in Fig. 290, thus limiting the current.

Increase of voltage above the striking point causes a further decrease in the impedance of the neon tube. Therefore more of the supply voltage will appear across the regulating resistance, while that across the neon tube and the load will tend to be unchanged

CHAPTER 7

THERMIONIC VALVES

THERMIONIC EMISSION

The phenomenon of current flow in an electrical conductor has been explained by the hypothesis that in a conductor, certain electrons in the outer orbits of the component atoms are comparatively loosely held to their parent nuclei, and that when an electrostatic field is superimposed on the conductor, a drift of the so-called "conductor electrons" results. This drift of electrons will be from the low potential to the high potential part of the electrostatic field, and will correspond to an electric current in the reverse direction.

When a metal is heated, the normal random motion of the conductor electrons is intensified, and at very high temperatures electrons will tend to leave the surface of the conductor altogether. This phenomenon is called "thermionic emission"—the name meaning simply the emission of electrons due to the application of heat—and on this phenomenon depends the operation of thermionic valves. A valve consists of a suitably heated "cathode", or emitter of electrons, together with one or more electrodes for collecting the electrons emitted, and for controlling and utilising the resulting flow of electrons.

In order to use a heated metal as a source of electrons, the cathode must be enclosed in an evacuated envelope. This is necessary for two reasons: firstly, most metals oxidise rapidly when heated to a high temperature in air; and secondly, if air (or any other gas) were present in the envelope, the emitted electrons, having attained sufficient velocity to leave the cathode, would collide with the molecules of the gas, causing ionisation of the gas and producing undesired results.

The cathode

The cathode may either be "directly heated", by constructing it in the form of a wire and passing an electric current through it, in which case it is called a "filament"; or it may be "indirectly heated", by making it in the form of a cylinder round a "heater wire" through which the heating current is passed. The heater of an indirectly heated valve is electrically insulated from the cathode, but is so attached to it mechanically that almost all the heat generated by the heater passes to the cathode. To permit this

passage of heat, the insulation between heater and cathode must be chosen and designed with its thermal behaviour as the primary consideration, and it is usually weak electrically. For this reason, care must be taken not to allow too high a potential difference to be developed between heater and cathode, or the insulation will break down. The maximum permissible heater-cathode voltage for most small valves is usually of the order of 100 volts.

Indirectly heated valves have three main advantages over the directly heated type. Firstly, there is a thermal reservoir effect between heater and cathode, due to the high thermal capacity of the insulation, so that the cathode remains at a constant temperature even when AC is used for the heater, whereas the temperature of the filament of a directly heated valve heated by AC varies between wide limits at twice the AC supply frequency. Owing to its thermal capacity, the cathode may take several minutes to reach its final working temperature, though it usually approaches this temperature after about 30 seconds. Secondly, since the cathode is electrically isolated from the heater, greater flexibility in circuit design is possible; in particular, several valves can be operated from the same heater supply, and yet have their cathodes connected to any desired points in their respective circuits without mutual interaction. Thirdly, since no heating current is passed through the cathode itself, the whole of the latter is at the same potential.

Cathode construction

The metal originally used for directly heated cathodes was tungsten, either pure or containing about one per cent. of thorium oxide, but the temperatures required were high, so also was the heating current. Most cathodes in use to-day (except in high-power radio transmitting valves) are of the oxide coated type; these give emission of electrons at dull red heat, and are economical in supply power. Directly heated filaments of this type consist simply of a wire of nickel, tungsten or nickel alloy coated with a preparation of barium, strontium or calcium oxides. The indirectly heated cathodes consist of a nickel tube coated with oxide forming the cathode, and inside this tube, and insulated from it, is a stout tungsten wire that forms the heater.

THE DIODE

The subsequent motion of the free electrons that surround the cathode as the result of thermionic emission may be influenced by electrostatic fields applied by means of further " electrodes ". Thermionic valves are classified according to the number of electrodes they possess, the simplest being the two-electrode valve or " diode ".

The diode consists of a cathode (either directly or indirectly heated) surrounded by a second electrode, the anode (or plate). Fig. 291 shows the assembly of the two types of diode, and the conventional symbol associated with each. Fig. 291*a* shows a

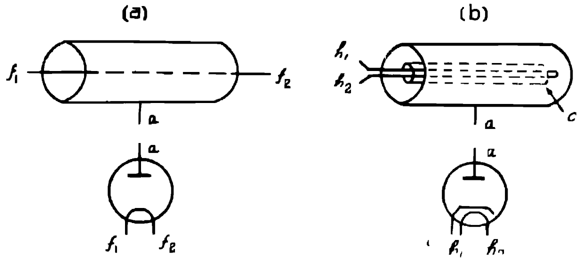


FIG. 291.—Arrangement of electrodes in a diode: (a) directly heated; (b) indirectly heated.

directly heated valve with a filament f_1f_2 and an anode a, whilst Fig. 291*b* shows an indirectly heated valve having a heater h_1h_2, a cathode c, and an anode a.

Space-charge

The rate of emission of electrons into the free space surrounding the cathode may be considered to be constant since it depends principally on the temperature of the cathode. What happens to these electrons afterwards depends on the electrostatic potential of the anode relative to the cathode. If the anode is made negative with respect to the cathode, the emitted electrons will be repelled

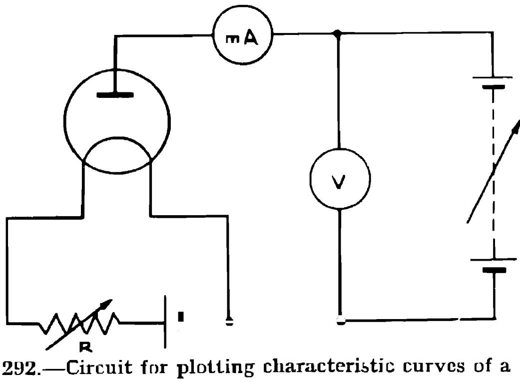


FIG. 292.—Circuit for plotting characteristic curves of a diode.

back into the cathode, and no current will flow between anode and cathode. If the anode is maintained at a positive potential relative to the cathode (Fig. 292), emitted electrons will be attracted towards the anode. The rate of arrival of electrons at the anode, that is, the anode or plate current, is, however, limited by the negative "space-charge" produced by the electrons in transit between cathode and anode. The number of electrons in transit at any instant is just sufficient to produce a negative space-charge which neutralises the attraction of the anode on the electrons just about to leave the cathode. All electrons emitted in excess of this number are at once repelled back into the cathode. Where the anode

current is limited by space-charge, it will be dependent on anode potential, and is substantially independent of the rate of emission of electrons by the cathode. Fig. 293 shows graphically the relationship between anode potential and anode current for a diode, and Fig. 294 shows the effect of the space-charge on the potential gradient in the valve.

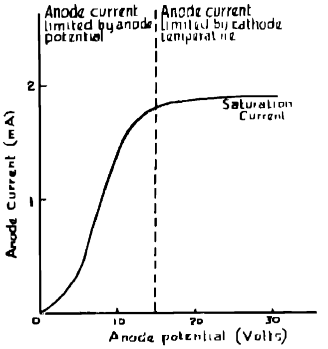


FIG. 293 Relationship between anode potential and anode current for a diode

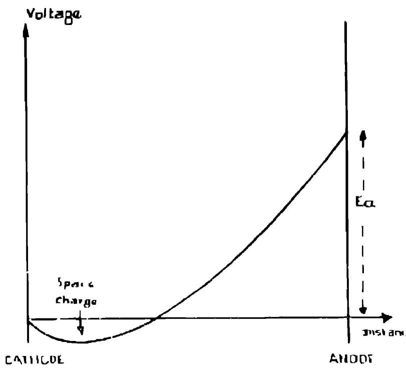


FIG. 294 Potential gradient in a diode, showing effect of space charge

The lower part of the curve in Fig. 293, corresponding to space-charge limitation of the anode current, can be represented by a three-halves power law, *i.e.* by :

$$I_a = k \, . \, E_a^{\frac{3}{2}} \tag{1}$$

where k is a constant depending on the construction of the valve.

As the anode potential is raised, a point is eventually reached where the space-charge effect produced by all the electrons emitted is not sufficient to balance the attraction due to the anode, and the anode current will be largely independent of the anode voltage, but will be determined by the rate of emission of electrons from the cathode and therefore by the temperature of the cathode.

The upper portion of the curve of Fig. 293 shows this condition.

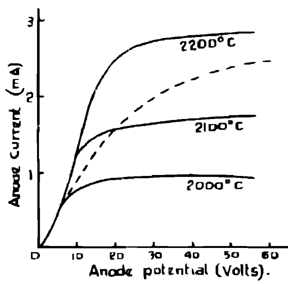


FIG. 295.—Effect of cathode temperature on characteristic curves of a diode

The limiting value of the anode current (that is, the value above which it is impossible to increase the anode current by increasing the anode voltage) is called the "saturation current", and the valve, in this condition, is said to be "saturated".

Fig. 295 shows how the saturation current is increased by raising the temperature of the cathode (by increasing the heater current). The curve of Fig. 293 and the solid curves of Fig. 295 are plotted for a tungsten emitter; thoriated tungsten gives curves of the same general shape, but in the case of oxide coated cathodes, saturation takes place much more gradually, as shown by the dotted curve of Fig. 295.

It has been stated that the electrons emitted from a heated

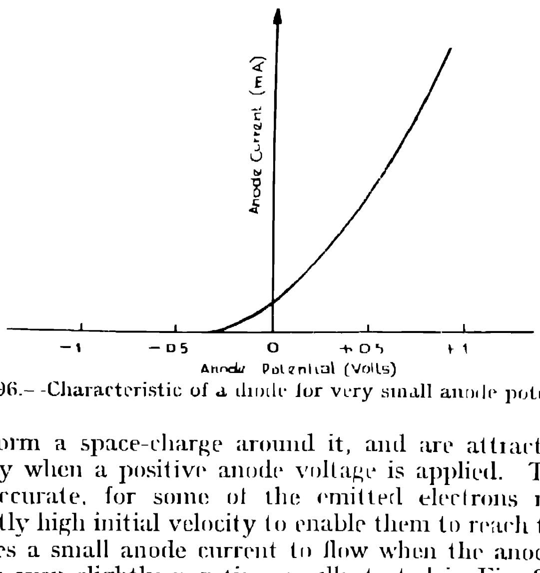


FIG. 296.—Characteristic of a diode for very small anode potentials.

cathode form a space-charge around it, and are attracted to the anode only when a positive anode voltage is applied. This is not strictly accurate, for some of the emitted electrons may have a sufficiently high initial velocity to enable them to reach the anode. This causes a small anode current to flow when the anode voltage is zero, or even slightly negative, as illustrated in Fig. 296. This current ceases to flow when the anode potential is made slightly more negative—the negative voltage on the anode necessary to reduce the current to zero being normally less than 1 volt.

Rectification using a diode

Fig. 297 shows the characteristic curve of a diode plotted for negative as well as positive anode potentials. It will be seen from this curve that the diode is a non-linear impedance of the same type as the metal rectifiers discussed in Chapter 6, except that the diode current falls to zero and does not increase in the reverse direction when a reverse voltage is applied.

Fig. 298 shows a diode used in a half-wave rectifier circuit. The operation of this circuit should be perfectly clear if it is

remembered that the diode will pass current from anode to cathode, but not from cathode to anode. It will be noted that in this circuit, an AC supply is used for a directly heated cathode; this is quite usual in diodes used for power supplies, but is not often employed

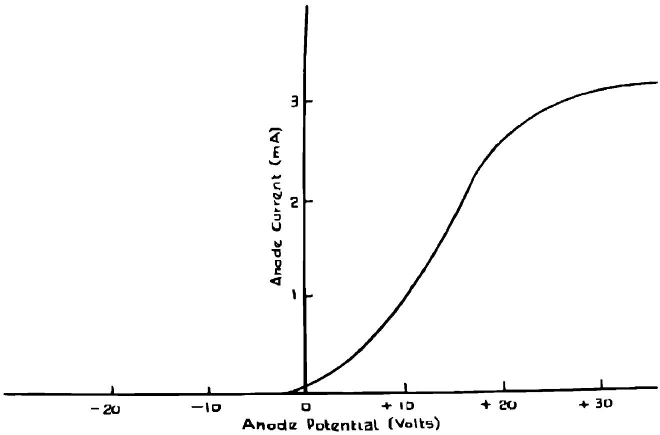


FIG. 297.—Characteristic of a diode for negative and positive anode potentials.

in the case of directly heated valves used for other purposes because it tends to give rise to excessive mains hum.

Fig. 299 shows a double-diode used in a full-wave rectifier circuit. A double-diode is a pair of diodes contained in the same envelope. In this example, the two diodes are sharing a common cathode, which is directly heated.

Fig. 300 shows a pair of diodes used in a full-wave rectifying and voltage-doubling circuit. The diodes shown here are indirectly heated, and it is easily seen that the two cathodes are at different potentials. If, therefore, the two valves are combined into a double-diode, separate cathodes will be necessary. If directly heated diodes are used for this circuit, a separate filament supply must, for the same reason, be used for each one.

THE TRIODE

The diode, which is the simplest form of thermionic valve, is limited to a single function, namely rectification. In the " triode ", the flow of electrons from cathode to anode is controlled by means of an additional electrode interposed between the cathode and the anode (*see* Fig. 301). This electrode, which is called the " grid " on account of the form taken in early examples of such valves, is in the form of an open mesh. The grid is normally operated at a negative potential relative to the cathode, and so it attracts no electrons to itself, so that no grid current flows; but it tends to repel those electrons which are being attracted to the anode.

The number of electrons reaching the anode is determined mainly by the electrostatic field near the cathode, and hardly

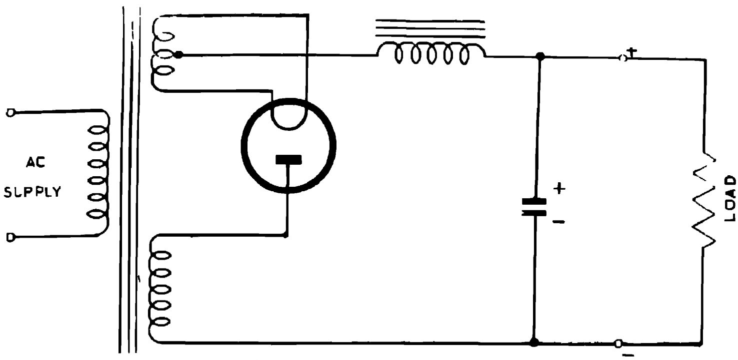


FIG. 298.—Half-wave rectification using a diode.

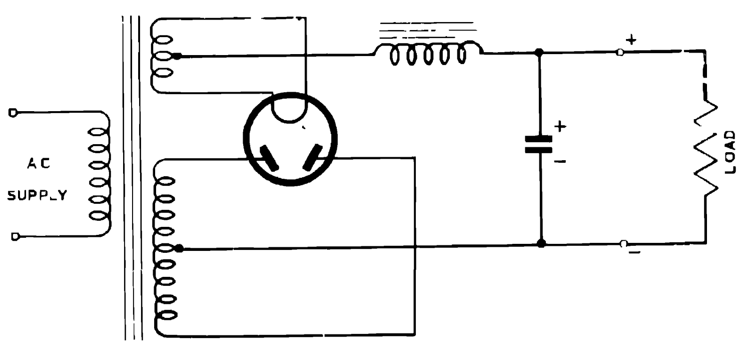


FIG. 299.—Full-wave rectification using a double-diode.

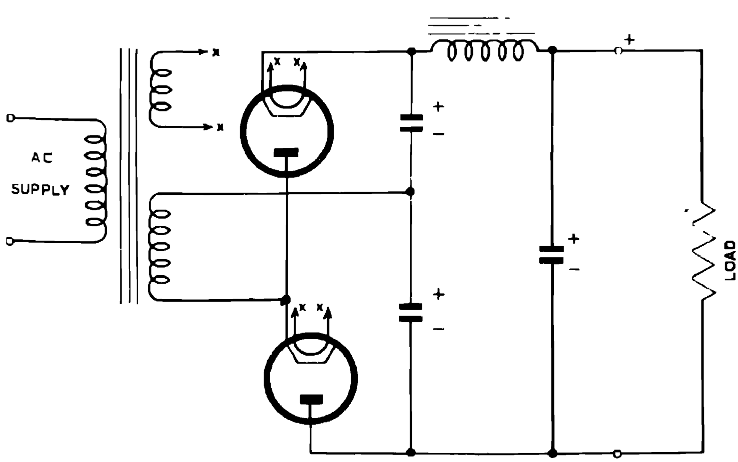


FIG. 300.—Full-wave rectification with voltage doubling using two diodes.

at all by the field in the rest of the space between the cathode and anode; for, near the cathode, the electrons are travelling slowly compared with those which have already moved some distance towards the anode, and the electron density in the inter-electrode space will be high near the cathode, decreasing

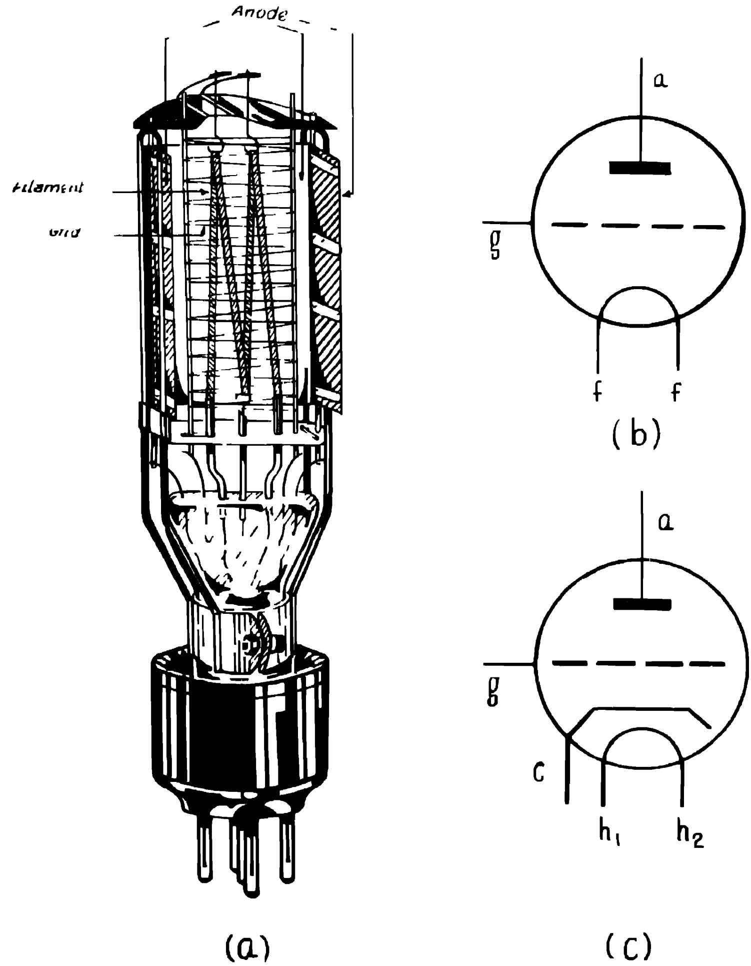


FIG. 301.—Arrangement of electrodes in a triode.

towards the anode. The total space-charge will be concentrated near the cathode, since once an electron has left this region it contributes to the space-charge for only a very brief interval of time. Thus the space current in a triode is determined by the

electrostatic field near the cathode produced by the combined effect of anode and grid potentials.

For a symmetrical grid structure it can be shown that the electrostatic field near the cathode is proportional to $\left(E_g + \frac{E_a}{\mu}\right)$ where E_g and E_a are the grid and anode potentials respectively and μ is a constant determined by the geometry of the valve. The total space current I_s varies with $\left(E_g + \frac{E_a}{\mu}\right)$ in exactly the same way that anode current varies with anode voltage for the diode; that is, over the lower part of the curve (for small values of I_s) :—

$$I_s = K\left(E_g + \frac{E_a}{\mu}\right)^{\frac{3}{2}} \quad (2)$$

K being a constant, determined by the dimensions of the valve, and $\left(E_g + \frac{E_a}{\mu}\right)$ being positive.

In general I_s is the sum of the anode current I_a and the grid current I_g, but if the grid is maintained at a sufficiently negative potential with respect to cathode, the grid current is zero since no electrons emitted by the cathode will go to the grid.

Thus for negative grid potential and $\left(E_g + \frac{E_a}{\mu}\right)$ positive :—

$$I_a = K\left(E_g + \frac{E_a}{\mu}\right)^{\frac{3}{2}} \quad (3)$$

for small values of I_a.

Under the same conditions, but with $\left(E_g + \frac{E_a}{\mu}\right)$ negative, I_a will be zero, since the electrostatic field near the cathode will then be such as to repel all the emitted electrons back into the cathode.

VALVE CONSTANTS AND CHARACTERISTIC CURVES

Equation 3 is a mathematical expression connecting I_a, E_g and E_a. It is always convenient to interpret such a result in a graphical form, but here, since there are three variables, this could only be done completely by plotting a series of points in three dimensions using three axes, mutually at right angles, for I_a, E_g and E_a. These points could then be joined to form a surface, the "characteristic surface" of the triode. Such a surface is drawn in perspective in Fig. 302.

Instead of plotting such a surface, certain characteristic curves may be derived; these are of two main types, the mutual and the anode characteristics.

Mutual characteristics

A surface such as that shown in Fig. 302 gives a complete picture of the interconnection between I_a, E_g and E_a; that is to say

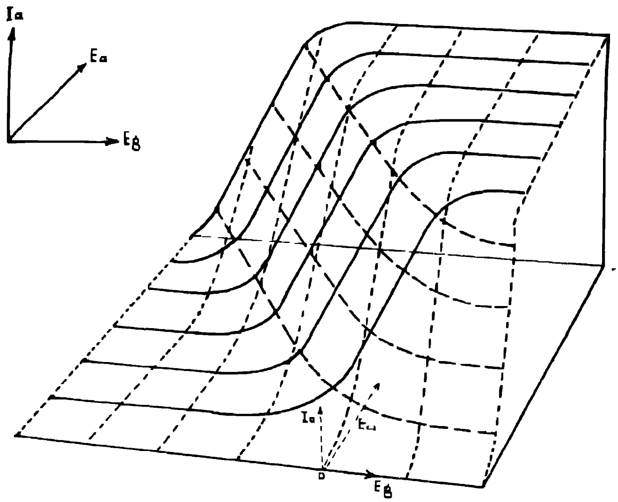


FIG. 302.—Characteristic surface for a triode

on such a surface all three quantities are continuously variable. Now suppose instead that a spot value of E_a, say $E_a = 50$ volts, is taken, and, with E_a kept constant at this value, the variation of I_a with E_g is plotted. The resultant two-dimensional curve will be a cross-section of the characteristic surface taken perpendicular to the axis of E_a through the point $E_a = 50$ volts. Similarly, curves may be plotted showing the connection between I_a and E_g for other values of E_a. The result is a family of mutual characteristics as shown in Fig. 303.

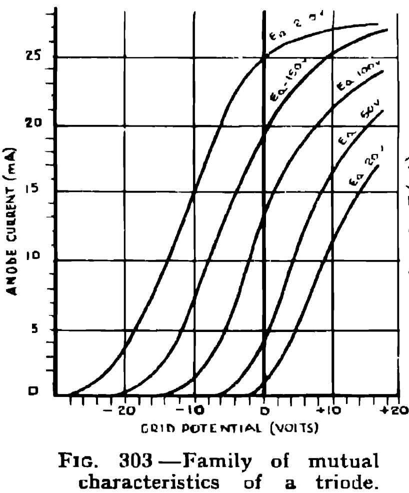


FIG. 303.—Family of mutual characteristics of a triode.

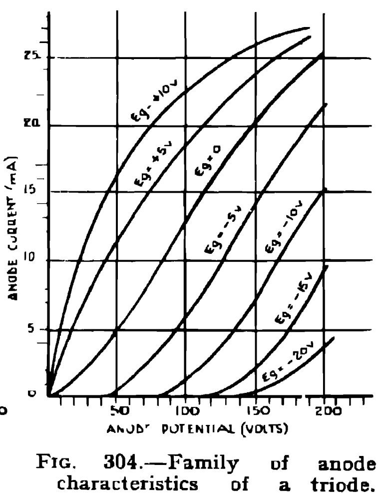


FIG. 304.—Family of anode characteristics of a triode.

This family of curves is equivalent to Fig. 302 except that it gives a picture that is discontinuous as far as E_a is concerned; it can never give a complete representation, but the picture may be made as detailed as required by interpolating for E_a between the curves. It will be noted that the curves have been extended into the region of positive grid potential. When the grid is slightly positive it will, owing to its open construction, attract only a negligible proportion of the total number of electrons emitted from the cathode, and the anode current will still be substantially equal to the total space current. A similarity is seen between the shape of these mutual characteristics and the shape of the diode characteristic shown in Fig. 293. The flattening of the curves for higher values of anode current is in both cases due to the saturation effect; that is, the total emission current fixes an upper limit for the anode current. The point on a mutual characteristic where the anode current becomes zero is called "cut-off", and the corresponding value of E_g is the "cut-off bias". The value E_{co} of the cut-off bias is clearly seen from equation 3 to be $E_{co} = -\frac{E_a}{\mu}$.

Anode characteristics

Another family of curves, the anode characteristics, may be derived by choosing fixed values of E_g, and plotting the variations of I_a with E_a. The result will be another set of cross-sections of Fig. 302, this time taken perpendicular to the axis of E_g.

This family of anode characteristics (Fig. 304) is also equivalent to Fig. 302, but this time the discontinuities occur in the values of E_g, and if a more detailed picture is required, it is necessary to interpolate for values of E_g between these curves. In the case of Fig. 304, the curves have not been drawn for sufficiently high values of anode current and anode voltage to show the effects of saturation, but these are similar to the case of the mutual characteristics; namely, that the anode current has an upper limit determined by the emission current, and all the curves will eventually flatten out at this value.

Amplification factor (μ)

The symbol μ has already been introduced as a constant depending solely on the geometry of the valve, and such that the electrostatic field near the cathode is proportional to $\left(E_g + \frac{E_a}{\mu}\right)$

It is in fact defined as the ratio of the relative effectiveness of the grid and anode voltages in producing electrostatic fields at the surface of the cathode, or (what clearly comes to the same thing) in producing anode current over that range where the anode current is limited by the space charge. In the practical case, due to slight asymmetry of the grid structure and to the necessity of having supporting wires, μ can no longer be considered a pure geometrical constant, but it will vary slightly with anode and grid voltages

and with the emission current. Nevertheless it is defined as follows :—

Amplification factor μ is the ratio of the relative effectiveness **of the grid voltage, to that of the anode voltage, in controlling the anode current.**

$$\mu = \frac{\dfrac{\partial I_a}{\partial E_g}}{\dfrac{\partial I_a}{\partial E_a}}$$

That the amplification factor μ defined in this way is equivalent to the μ of equations 2 and 3 on page 331 may be seen by partial differentiation of equation 3. Thus :—

$$\frac{\partial I_a}{\partial E_g} = \frac{3}{2}K\left(E_g + \frac{E_a}{\mu}\right)^{\frac{1}{2}}$$

and
$$\frac{\partial I_a}{\partial E_a} = \frac{3}{2}K\left(E_g + \frac{E_a}{\mu}\right)^{\frac{1}{2}} \cdot \frac{1}{\mu}$$

giving again :—
$$\mu = \frac{\dfrac{\partial I_a}{\partial E_g}}{\dfrac{\partial I_a}{\partial E_a}}$$

To avoid using partial derivatives, the amplification factor μ can be defined in terms of the voltage increments, δE_g and δE_a, in grid and anode potentials respectively, that keep the plate current constant. Thus an increase δE_a in anode potential gives an increase δI_a in anode current, with grid potential constant at E_g ; and it requires an increase $-\delta E_g$ (a decrease) in grid potential to restore the anode current to its original value, with anode potential constant at $E_a + \delta E_a$.

In this case the amplification factor is defined as :—

$$\mu = \frac{\delta E_a}{\delta E_g} \tag{4}$$

It has been stated that in the practical valve, μ will depend very slightly on the operating conditions due to asymmetry of the grid, but for the most part it may be taken as a constant except for a tendency to become lower as cut-off is approached.

Mutual conductance g (sometimes represented by g_m, G_m, or S)

The mutual conductance g (sometimes referred to as transconductance) is defined as the rate of change of plate current with respect to change of grid potential.

i.e.
$$g = \frac{\partial I_a}{\partial E}$$

Or, using the approximate notation of finite increments, if an increment δE_g in grid potential causes an increase δI_a in anode current, when the anode voltage is kept constant,

$$g = \frac{\delta I_a}{\delta E_g} \tag{5}$$

Mutual conductance has the dimensions of a conductance and is expressed either in micromhos (μmho) or more commonly in milliamps per volt. It will be seen from equation 5 that g is the gradient or slope of the mutual characteristic. Since the mutual characteristics (*see* Fig. 303) are approximately straight and parallel over most of their length, g is substantially independent of operating conditions, except for a decrease near cut-off. In certain valves, however, the grid is made definitely asymmetrical, with a view to causing a large variation in mutual conductance under different operating conditions. Such a valve is called a "variable-mu" valve and will be discussed later (p. 360).

AC resistance R_a (or ρ)

The anode characteristic resistance (also called the internal resistance or the internal impedance) is defined as the rate of change of anode potential with respect to change in anode current.

$$R_a = \frac{\partial E_a}{\partial I_a}$$

Or, if an increment δE_a in anode potential causes an increase δI_a in anode current, when the grid potential is kept constant,

$$R_a = \frac{\delta E_a}{\delta I_a} \tag{6}$$

R_a has the dimensions of a pure resistance and is measured in ohms (or megohms). It is the reciprocal of the slope of the anode characteristic, and where the anode characteristics are straight and parallel it will be constant (*see* Fig. 304).

It is important to notice that the AC resistance is concerned only with the ratio of *changes* of anode potential to *changes* in anode current, and *not* with the ratio of *total* anode potential to *total* anode current. This is also true of g and μ.

Reference is frequently made to "low impedance" and "high impedance" valves. A low impedance valve is a valve having a small value of R_a, whereas a high impedance valve has a high value of R_a.

Relation between μ, g and R_a

The three valve constants μ, g and R_a are not independent, but as may be seen from equations 4, 5 and 6, they are connected by the relation :—

$$\mu = R_a \times g \tag{7}$$

where g is in mhos, and not in mA/V.

Derivation of valve constants from characteristics

The valve constants μ, g and R_a may all be derived under given operating conditions from a study of *either* a set of mutual characteristics *or* a set of anode characteristics.

Fig. 305 shows a set of mutual characteristic curves for a CV 1664 (AR13), a low-impedance triode.

Suppose the operating conditions are given as $E_a = 130$ volts, $E_g = -4 \cdot 5$ volts. From the curve corresponding to $E_a = 130$ volts it can be seen that $E_g = -4 \cdot 5$ volts gives $I_a = 6 \cdot 5$ mA (point P, Fig. 305).

Now keeping E_a constant at 130 volts (that is, remaining on the same curve), it can be seen that a change of $1 \cdot 5$ volt in grid potential to -3 volts gives an increase in anode current to $10 \cdot 25$ mA (point Q, Fig. 305).

i.e. $\delta E_g = 1 \cdot 5$ volt, and $\delta I_a = 3 \cdot 75$ mA

Therefore $g = 2 \cdot 5$ mA/volt

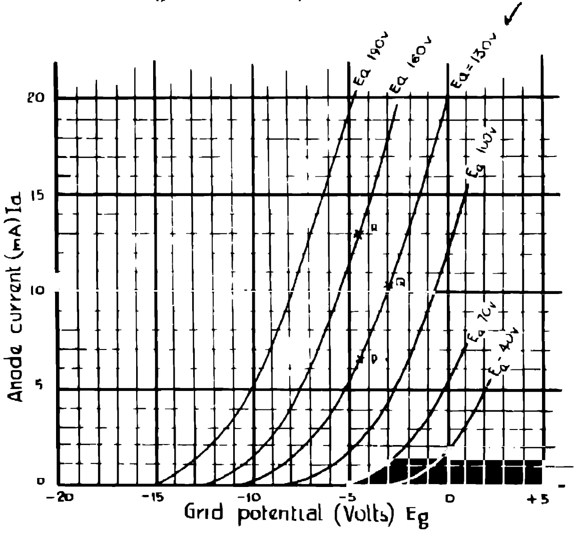


FIG. 305.—Mutual characteristics of a triode (CV 1664).

Now going back to the initial operating point, *i.e.* $E_a = 130$ volts, $E_g = -4 \cdot 5$ volts, $I_a = 6 \cdot 5$ mA, and keeping E_g constant at $-4 \cdot 5$ volts, a change in E_a involves leaving the curve corresponding to $E_a = 130$ volts; but considering the curve corresponding to $E_a = 160$ volts, it can be seen that the anode current for $E_a = 160$ volts and $E_g = -4 \cdot 5$ volts is 13 mA (point R), so that an increase of 30 volts in E_a gives an increase of $6 \cdot 5$ mA in I_a provided E_g is constant at $-4 \cdot 5$ volts.

i.e. $\delta E_a = 30$ volts, $\delta I_a = 6 \cdot 5$ mA

Thus $$R_a = \frac{30}{6 \cdot 5 \times 10^{-3}} = 4600\Omega$$

Since $$\mu = R_a \times g$$

$$\mu = 4600 \times \frac{2 \cdot 5}{1000} \text{ (} g \text{ reduced to amps/volt, } i.e. \text{ mhos)}$$

i.e. $$\mu = 11 \cdot 5$$

Now consider Fig. 306, which shows a set of anode characteristics for the same valve. Suppose the operating point is given as $E_g = -4\cdot5$ volts, $E_a = 130$ volts. From the curves, $I_a = 6\cdot5$ mA (point S).

Now keeping E_g constant, an increase in E_a from 130 volts to 160 volts gives an increase in I_a from 6·5 mA to 13·0 mA (point T).

i.e. $\delta E_a = 30$ volts $\qquad \delta I_a = 6\cdot5$ mA

Thus $$R_a = \frac{30}{6\cdot5 \times 10^{-3}} = 4600 \text{ ohms.}$$

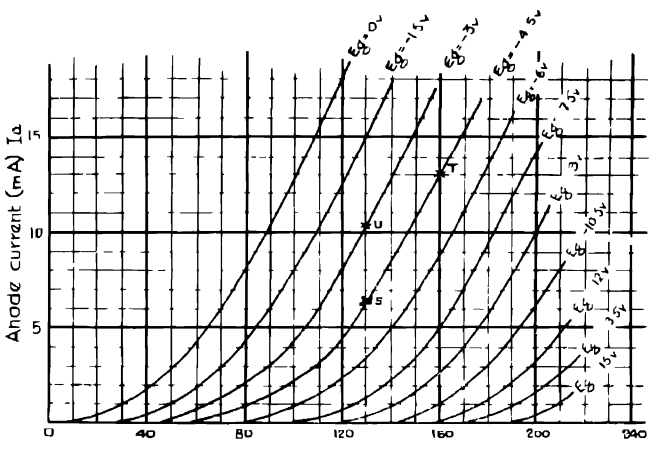


FIG. 306.—Anode characteristics of a triode (CV 1664).

Now taking the same operating point and keeping $E_a = 130$ volts, an increase in E_g from $-4\cdot5$ volts to -3 volts gives an increase in I_a from 6·5 mA to 10·25 mA (point U).

i.e. $\delta E_g = 1\cdot5$ volts $\qquad \delta I_a = 3\cdot75$ mA

Therefore $$g = \frac{3\cdot75}{1\cdot5} = 2\cdot5 \text{ mA/volt}$$

and since $$\mu = R_a \times g$$

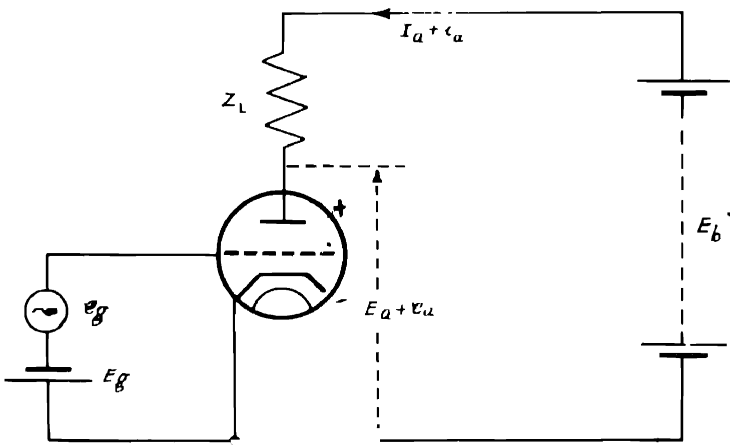
$$\mu = 4600 \times \frac{2\cdot5}{1000}$$

i.e. $$\mu = 11\cdot5$$

This example emphasises once more that the mutual characteristics and anode characteristics are merely two different ways of imparting exactly the same information. One or other of these sets of characteristics, sometimes both, is given in valve data sheets in order that suitable operating conditions may be chosen to suit the purpose for which the valve is to be used. Even if only one set of characteristics is available, it contains all the information required.

THE TRIODE VALVE AS AN AMPLIFIER

Fig. 307 shows a triode arranged as an amplifier in the simplest possible way. The signal to be amplified is applied to the grid in the form of an alternating voltage. E_g is a steady voltage applied in series with the signal, of such a magnitude as to ensure that the grid is always negative with respect to the cathode. The anode is maintained at a high positive potential relative to the cathode by a battery (the high-tension or HT battery, of voltage E_b) acting in series with the load impedance Z_L. Anode current will flow, and there will be a potential drop across the load impedance Z_L; the potential of the anode relative to cathode will therefore be less than E_b.



$$E_a + e_a = E_b - (I_a + i_a) Z_L$$
$$E_a = E_b - I_a Z_L$$
$$e_a = -i_a Z_L$$

FIG. 307.—Simple circuit of a triode amplifier stage.

When no signal is being applied to the grid ($e_g = 0$), the DC bias voltage E_g will determine the value of the steady anode current I_a and the steady anode voltage E_a. If the small alternating voltage e_g is now applied in the grid circuit, an alternating component i_a is produced in the anode current; this will give an alternating voltage drop across Z_L, resulting in an alternating component e_a in the anode voltage. This alternating component may be utilised as required.

Amplifiers are subdivided into two classes according to the use made of the voltage developed across the load impedance. In the simplest case, it is required merely to develop the maximum possible voltage across the load, and to apply it to the grid of a second valve

for further amplification. Since the grid circuit of a valve is voltage-operated without consumption of power, it is immaterial what *power* is developed in Z_L. Such a stage of amplification is called a "voltage amplifier".

In certain cases, however, the main consideration is the *power* developed in the anode load. For example, in an audio amplifier the *loudness* of the audible result will depend on the *power* dissipated in the loud-speaker or telephone receiver which will be acting as the anode load of the final stage of amplification. This last stage, in which power is the primary concern, is called a "power amplifier", and amplifiers of this type require separate treatment which will be given later.

Dynamic characteristics

In the previous section (*see* Figs. 305 and 306) an operating point on the characteristics was chosen, determined by the values of E_a and E_g, and the alterations in anode current produced by an additional voltage e_g applied to the grid were then considered. This will now be considered in greater detail.

The mutual characteristics so far considered have shown how I_a varies with E_g provided E_a is kept *constant*. Similarly, in the case of the anode characteristics, there was the proviso that E_g be kept *constant*. These characteristics are known as the "static characteristics", and give certain information about the valve itself, making possible the choice of suitable valves and suitable working conditions for any particular purpose.

But it may now be seen that if a valve is connected up in a particular circuit, as in Fig. 307, with an anode load of impedance Z_L, then if the potential on the grid is varied the potential on the anode is also varied. This does not, however, occur under the conditions under which the static mutual characteristics are plotted, *i.e.* constant E_a. When I_a changes, so does the voltage drop across Z_L, and since the anode voltage is applied from a battery of constant voltage E_b, the anode voltage E_a will change. Thus to get a true picture in this case a set of characteristics is required giving the variation of I_a with E_g, subject to the simultaneous and consequent variations of E_a, the extent of which will vary with the load impedance. Such a set of characteristics is called a set of "dynamic" mutual characteristics, and would be characteristic not of the valve itself, but of the valve when connected to a particular value of anode load. This would appear to necessitate a set of dynamic characteristics for every value of load impedance; fortunately the dynamic characteristics corresponding to any particular value of load impedance can be deduced from the static characteristics as will be shown by an example. For this reason, the latter type only will be found on valve data sheets.

Calculation of gain from dynamic characteristics

Consider the circuit of Fig. 307. The case where the available

HT voltage (E_b) is 190 volts will now be considered, supposing that the valve is a CV 1664 whose static mutual characteristics were shown in Fig. 305 and are repeated in Fig. 308. For simplicity, consider a purely resistive load of 30,000Ω.

Now when $I_a = 0$, there will be no potential drop across the anode load, and $E_a = 190$ volts. From the static characteristic corresponding to 190 volts, it is seen that $I_a = 0$ corresponds to $E_g = -15$ volts. The point A is therefore on the dynamic characteristic.

When $I_a = 3$ mA, the potential drop across the resistive load will be 90 volts, so that $E_a = 100$ volts ; the point B, corresponding to $E_a = 100$ volts and $I_a = 3$ mA, will therefore lie on the dynamic characteristic. In the same way by assuming other values for I_a

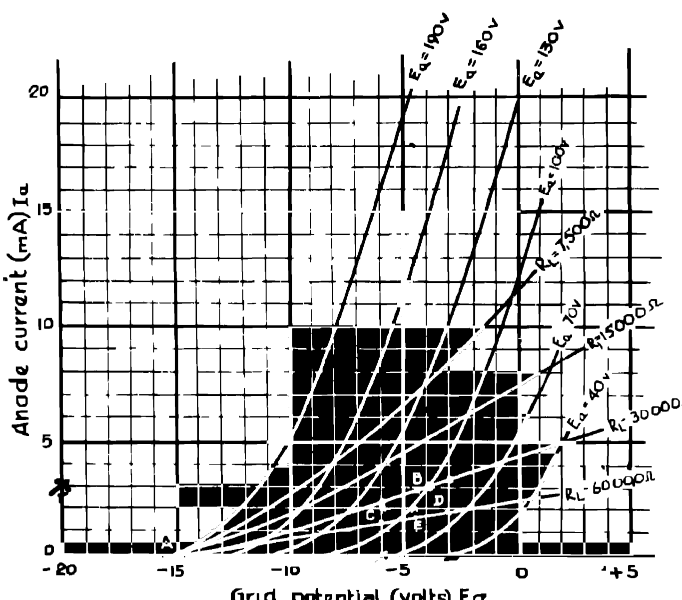


FIG. 308 —Dynamic mutual characteristics for various values of load resistance

the dynamic characteristic corresponding to a resistive load of 30,000 ohms (or for any other value of resistance) may be plotted completely. A number of these dynamic characteristics is shown in Fig. 308.

These facts are at once apparent :—

(*a*) The characteristics are for the most part very straight except for slight curvature near cut-off.

(*b*) The higher the load resistance the smaller the slope of the dynamic characteristics, and *vice versa*.

(*c*) The smaller the load resistance, the more nearly does the dynamic characteristic coincide with the static characteristic

$E_a = 190$ volts, and the greater is the curvature at the lower end.

The dynamic characteristic corresponding to an available HT supply of 190 volts and a load resistance of 30,000 ohms having been deduced, the operating point on that characteristic (say, $E_g = -4 \cdot 5$ volts, corresponding to an anode current of $2 \cdot 8$ mA) may be chosen. Now suppose an alternating voltage is applied to the grid in addition to the steady bias $-4 \cdot 5$ volts, and let the peak value of the signal be $1 \cdot 5$ volts; then the grid potential will vary between -6 volts and -3 volts, and it is now apparent that for the given load, $30{,}000\Omega$, the variations in anode current will all lie on the dynamic mutual characteristic corresponding to that value of load. Therefore I_a will vary between $2 \cdot 3$ mA and $3 \cdot 3$ mA; that is, a total variation of 1 mA, or $0 \cdot 5$ mA on either side of the "no signal" anode current of $2 \cdot 8$ mA. The change in anode current will be proportional to the change in grid voltage, and an undistorted signal will result provided that the dynamic characteristic is straight throughout the range of the variation of grid voltage. For this reason the operating point is chosen in the centre of the straight portion of the characteristic lying in the range of negative values of grid potential. This allows the maximum voltage signal to be applied to the grid without causing distortion; for, generally speaking, the voltage on the grid, relative to the cathode, must always be sufficiently negative to prevent the flow of grid current, and yet, on the other hand, not so negative as to cause operation over the lower curved portion of the dynamic characteristic.

With a signal of peak voltage $1 \cdot 5$ volts on the grid, an alternating anode current of peak value $0 \cdot 5$ mA flows in the load resistance of $30{,}000\Omega$, thus developing a peak voltage of 15 volts across the load. Thus in the particular circuit considered there is a voltage amplification of 10. This is known as the "stage gain", and it depends on the value of the load resistance.

Suppose the load is now changed to $60{,}000\Omega$. Choosing again $E_g = -4 \cdot 5$ volts, it is seen that a $1 \cdot 5$ volt peak signal will cause an alternating anode current of peak value $0 \cdot 275$ mA (points C and D, Fig. 308) about the steady value $1 \cdot 6$ mA (point E). This gives a peak alternating voltage of $16 \cdot 5$ volts across the $60{,}000\Omega$ load—that is, a stage gain of 11.

It will be verified later, that the higher the load resistance, the higher the stage gain; but the stage gain will always be less than the amplification factor (μ) of the valve as derived from its static characteristic. (For this valve, $\mu = 11 \cdot 5$.)

The load line

Corresponding to the dynamic mutual characteristics, are the "load lines" on the graph of the anode characteristics. Fig. 309 shows the static anode characteristics for the CV 1664 (the triode considered in the preceding paragraphs) together with load lines

for various values of resistive anode load, assuming again an available HT supply of 190 volts.

From Fig. 307 it can be seen that (taking $Z_L = R_L$, a pure resistance) :—

$$E_a = E_b - I_a R_L \tag{8}$$

If R_L and E_b are constants, the two variables I_a and E_a may be plotted in the form of a graph. This will be a straight line (since equation 8 is linear in E_a and I_a), and it will clearly pass through the points given (i) by $E_a = E_b$, $I_a = 0$ and (ii) by $E_a = 0$, $I_a = \dfrac{E_b}{R_L}$. This straight line is called the "load line" for the particular load

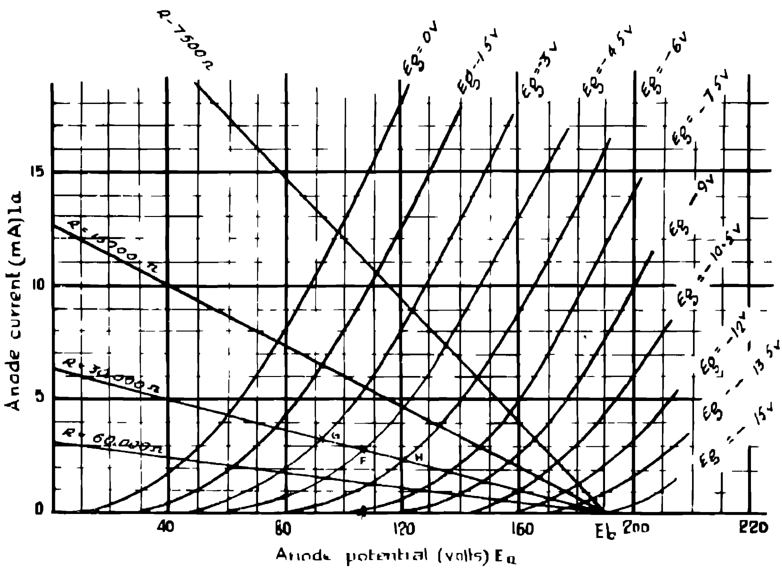


FIG. 309 —Anode characteristics showing load lines for various values of anode load.

considered and the HT supply available. Consider in particular the load line for $R_L = 30{,}000\Omega$ and $E_b = 190$ volts. Suppose that $E_g = -4{\cdot}5$ volts. The $-4{\cdot}5$ volts anode characteristic meets the $30{,}000\Omega$ load line in the point F, corresponding to $E_a = 106$ volts, $I_a = 2{\cdot}8$ mA. Let this be taken as the operating point; that is, let E_g be made $-4{\cdot}5$ volts; because the valve has a $30{,}000\Omega$ load and an available HT supply of 190 volts, this point F shows that the anode current will be $2{\cdot}8$ mA. Now suppose a signal of peak value $1{\cdot}5$ volts is applied to the grid, *i.e.* E_g will vary between -3 volts and -6 volts (points G and H). From the points of intersection of the corresponding characteristics with the $30{,}000\Omega$ load line, it can be seen that I_a varies between $3{\cdot}3$ mA and $2{\cdot}3$ mA and E_a between 91 and 121 volts. Thus for equal swings of grid

voltage about the standing bias, equal swings in the value of anode current (and of anode voltage) are obtained; this implies no distortion. In choosing a load, therefore, a value must be selected such that the corresponding load line makes equal intercepts on the anode characteristics. The selection of a load line making equal intercepts is exactly equivalent to selecting a straight dynamic characteristic.

Returning to the swings in anode current, these have a peak value of $\frac{3 \cdot 3 - 2 \cdot 3}{2}$, *i.e.* 0·5 mA, giving a peak value to the alternating voltage across the 30,000Ω load of 15 volts. Since the applied signal was 1·5 volts peak, this represents a stage gain of 10—which result has already been obtained by considering the dynamic mutual characteristic corresponding to a load of 30,000 ohms.

Thus, in the same way that the static mutual and anode characteristics are exactly equivalent as far as imparting information about the valve itself is concerned, so the dynamic mutual characteristic and the load line are equivalent ways of expressing the behaviour of the valve with a given resistive anode load, and either method may be used in designing a stage. It is, however, somewhat easier to detect inequality of intercepts on the load line than to detect slight curvature of the dynamic characteristic, so that the load line method of choosing operating conditions is the one usually employed.

Equivalent circuit

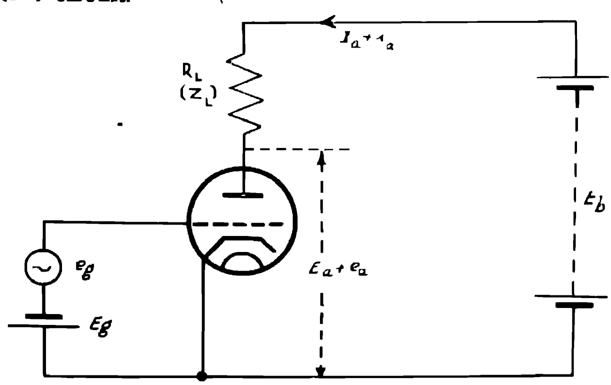


FIG. 310.—Simple circuit of a triode amplifier.

Consider the circuit of Fig. 310. It has already been demonstrated that the steady value of anode current I_a is determined by the steady values of grid voltage E_g and anode voltage E_a. Suppose now the grid voltage is changed by an amount e_g. If the anode voltage were kept constant, the corresponding change in anode current would be i_a' given by :—

$$i_a' = g\, e_g \tag{9}$$

It has been seen, however, that the anode voltage will not remain constant owing to the change in potential drop across R_L caused by the change in anode current.

Now suppose that the anode voltage changes by an amount e_a, due to this cause or any other; then the corresponding change i_a'' in anode current, assuming the grid voltage to be kept constant, is given by :—

$$i_a'' = \frac{e_a}{R_a} \tag{10}$$

Therefore the total variation in anode current due to simultaneous changes in anode voltage and grid voltage will, to the first order, be given by :—

$$i_a = g \cdot e_g + \frac{1}{R_a} e_a \tag{11}$$

But the change in anode voltage consequent upon a change e_g of grid potential will be :—

$$e_a = -\, i_a R_L \tag{12}$$

The negative sign occurs because a positive increment in grid voltage will cause an increase in voltage drop across R_L and therefore a *reduction* in anode voltage (*see* Fig. 311).

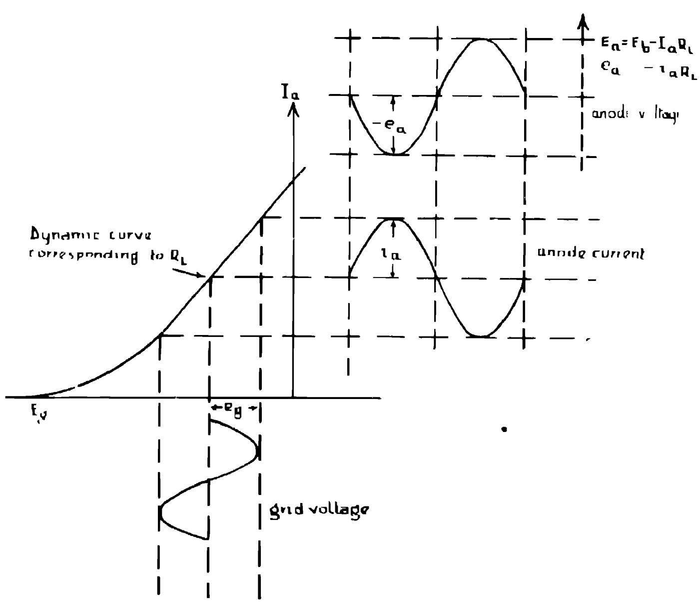


FIG. 311.—Illustrating relation between e_g, i_a and e_a for resistive load.

Eliminating e_a between equations 11 and 12 :—

$$i_a = g e_g - \frac{i_a R_L}{R_a}$$

i.e.

$$i_a = \frac{R_a g}{R_a + R_L} e_g$$

i.e.
$$i_a = \frac{\mu}{R_a + R_L} e_g \tag{13}$$

From equation 13, it can easily be seen that Fig. 312 is the equivalent circuit of Fig. 310. This, it is to be noted, applies only to AC components of current and voltage. The separation of AC and DC components can be justified by the superposition theorem (p. 231), as long as the linear part of the valve characteristic is

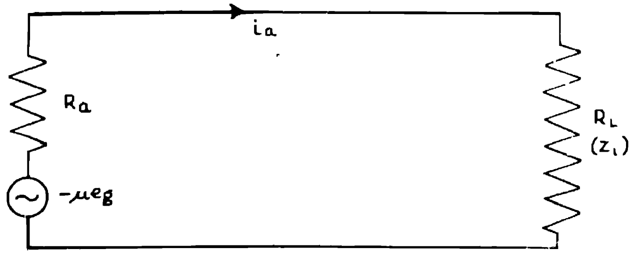


FIG. 312.—Equivalent circuit of triode amplifier (constant-voltage generator form).

used. It is usual to take i_a in a clockwise sense, and regard the generator as having a negative EMF. Equation 13, with the new interpretation on the direction of i_a, becomes :-

$$i_a = \frac{-\mu e_g}{R_a + R_L} \tag{14}$$

If R_L in Figs. 310 and 312 be replaced by a general impedance Z_L, equation 14 becomes :—

$$i_a = \frac{-\mu . e_g}{R_a + Z_L} \tag{15}$$

The change in voltage produced across Z_L will then be given by :—

$$e_L = \frac{-\mu e_g Z_L}{R_a + Z_L} \tag{16}$$

The stage gain M will be the ratio of the magnitudes of e_L and e_g.

$$\therefore \qquad M = \frac{\mu |Z_L|}{|R_a + Z_L|} \tag{17}$$

From this equation it will be seen that the stage gain M is always less than the amplification factor μ of the valve, but that it may be made as near to μ as required by taking a sufficiently large value of $|Z_L|$, provided that there is no limit to the value of E_b available. The equivalent circuit of Fig. 312 is known as the "constant-voltage generator" form of the equivalent circuit of an amplifier, and is the basis of most amplifier design calculations. This form is usually the most convenient for use when triodes are involved, but the "constant-current generator" form (Fig. 313) is preferable when R_a is much greater than the anode load Z_L, as is often the case with pentodes. The equivalence of the two forms may be verified as follows :—

Dividing numerator and denominator of the right-hand side of equation 15 by R_a:—

$$i_a = \frac{\frac{-\mu}{R_a} e_g}{1 + \frac{Z_L}{R_a}}$$

i.e.

$$i_a = -g\, e_g \cdot \frac{R_a}{R_a + Z_L} \tag{18}$$

The voltage developed across the anode load is therefore

$$E_L = -g\, e_g \cdot \frac{R_a Z_L}{R_a + Z_L} \tag{19}$$

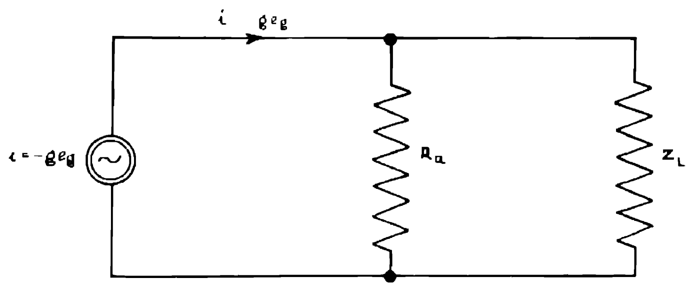


FIG. 313 Constant-current form of the equivalent circuit of an amplifier.

This is the voltage developed by a constant current $-g\, e_g$ across a circuit consisting of R_a and Z_L in parallel (*see* Fig. 313).

Hence
$$M = g \frac{R_a |Z_L|}{|R_a + Z_L|} \tag{20}$$

which is clearly equivalent to equation 17.

Grid current

The question of the flow of grid current has so far been dismissed because :—

(*a*) The grid has in general been kept so negative with respect to the cathode, that the grid will not collect electrons emitted by the cathode ; and

(*b*) Even if the grid is slightly positive with respect to the cathode, the grid current will generally be very small since the grid surface is very small.

Characteristics showing the grid current plotted against grid potential are often given on the mutual characteristic, using a magnified scale for current, since the grid current is of the order of microamps. The grid current will depend on anode potential, being small when the anode potential is high ; for very small anode voltages, the grid current may be of the order of milliamps for positive grid voltages approaching the anode potential.

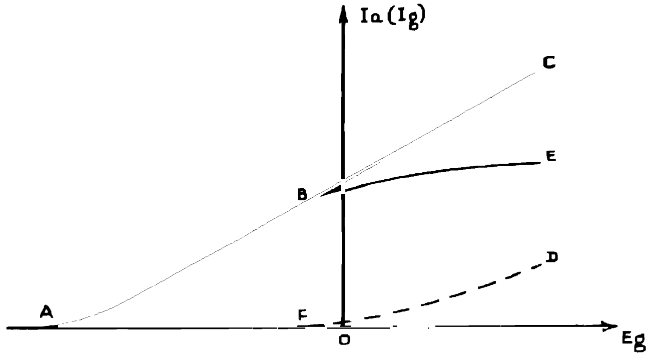


FIG. 314.—Effect of grid current on the dynamic mutual characteristic.

Although in certain instances it is desirable that grid current should flow (*e.g.*, provision of grid leak bias), it is in general undesirable in amplifiers for two reasons.

(*a*) Fig. 314 shows a dynamic mutual characteristic *ABC* such as would be obtained if no grid current flowed for positive grid voltages—that is to say, *ABC* is really a graph of total space current against grid voltage. If a portion of this space current is due to grid current, shown by the graph *FD*, then *ABE* will represent the graph of anode current against grid potential. Thus, due to the flow of grid current, a dynamic characteristic tends to take the form of the curve *ABE*. This curvature of the dynamic characteristic for positive grid potentials will introduce distortion, and the output will no longer correspond to the input. In the same way, flow of grid current will cause the anode characteristics to intersect the load line in unequal intercepts at positive grid voltages.

(*b*) A flow of grid current implies a dissipation of power in the grid circuit, and the valve is no longer strictly a voltage-operated device. Thus flow of grid current causes small power losses.

Effect of gas

If even a very small quantity of gas is present in a valve, the electrons emitted by the cathode will collide with the molecules of the gas, thereby ionising them. The positive ions thus formed are attracted towards the cathode and the negative control grid. The negative ions formed are attracted towards the anode. Those ions which bombard the cathode tend to destroy its electron-emitting properties, while those reaching the grid constitute a negative grid current, flowing from grid to cathode through the external circuit. This current will develop a voltage across the grid-cathode resistance in such a direction as to make the grid less negative, thereby increasing the space current; this increases the number of positive ions produced, causing an increase in grid

current and the application of a further positive voltage to the grid. If the grid-cathode resistance is sufficiently high, this process may become cumulative, and the space current rises rapidly to an excessive value, causing damage to the valve.

Effect of secondary emission

A further danger to valves caused by a grid-cathode circuit of high resistance lies in the possibility of "secondary emission" from the grid. If one primary electron from the cathode, impinging on the grid, causes the emission from the grid of several secondary electrons, the grid will acquire a positive charge; and if the grid-cathode resistance is high, this charge will not leak away immediately. More primary electrons will therefore be attracted to the grid, causing the emission of further secondary electrons, an increase in the positive charge on the grid, and a corresponding increase in the space current of the valve. This process is cumulative, and the space current may increase to an excessive value and cause damage to the valve.

Effects of open-circuited ("free") grid

When a valve is operated with a circuit of unduly high resistance between grid and cathode, one of three things may happen :—

(1) If any gas is present in the valve, positive ions reaching the grid may cause the space current to rise rapidly to a dangerous value.

(2) If secondary emission takes place from the grid, the space current may again rise rapidly to a dangerous value.

(3) If no gas is present, and no secondary emission takes place from the grid, electrons leaving the cathode may accumulate on the grid, thus imparting to it a negative charge. A condition of equilibrium is soon reached in which the grid has a large negative charge, and the space current is reduced almost to cut-off. Although in this case there is little possibility of damage being done, the valve will clearly cease to perform its normal function.

These three ill-effects can be avoided by keeping the resistance of the grid-cathode circuit low; and valve manufacturers, in their tables of operating data, usually state the maximum value of this resistance for each type of valve. For most small receiving valves, it is of the order of 2MΩ, while for power valves (in which the rise of current due to gas or positive ions is much more serious) it is much less.

Interelectrode capacity

Other characteristics of the valve that must be considered are the capacities that exist between the various pairs of electrodes.

In the case of a triode there are three such capacities; these are shown in Fig. 315 :—

(*a*) The *grid-cathode capacity*, C_{gc}, effectively in parallel with the input circuit and sometimes called the ***input capacity.***

(*b*) The *anode-cathode capacity*, C_{ac}, in parallel with R_L and sometimes called the *output capacity*.

(*c*) The *grid-anode capacity*, C_{ga}, providing a leakage path by which energy may be transferred from anode circuit to grid circuit, and therefore sometimes referred to as the *leakage capacity*.

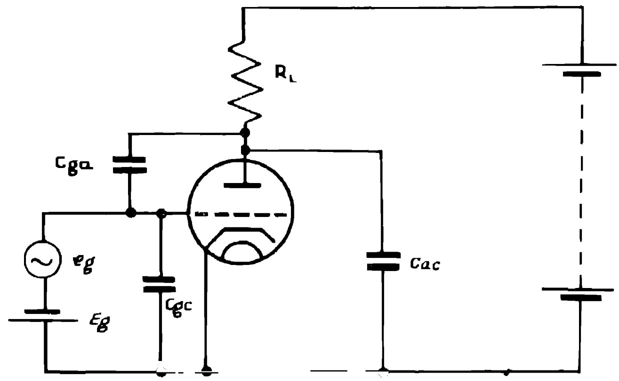


FIG. 315.—Triode amplifier showing interelectrode capacities.

Each of these capacities has in effect a high resistance shunt across it due to DC leakage resistance between electrodes, but this resistance is in general very high and will be neglected, as it introduces unnecessary complications into the equivalent circuit. The effect of these capacities is as follows : —

Grid-cathode capacity gives the valve a finite input impedance ; this causes a current to flow in the grid circuit, and so causes dissipation of power in the resistive component of the grid circuit impedance.

Anode-cathode capacity puts a shunt across the anode load and reduces the stage gain at high frequencies.

Grid-anode capacity.- This is the most important of the interelectrode capacities, and will be discussed in greater detail. Clearly it provides, particularly at high frequencies, a leakage path between anode and grid circuits by means of which voltage changes on the anode will be fed back to the grid. If this voltage fed back to the grid is of sufficient magnitude and is in the correct phase, the valve may cease to function as an amplifier and become an oscillator. This leakage capacity is the cause of instability in the triode, particularly when used as an amplifier at radio frequencies. Historically, the evolution of the screen–grid and pentode valves was a direct consequence of attempts to reduce the grid-anode capacity to such an extent that the valve would be stable as an RF amplifier ; but these pentodes have other advantages over triodes, such as very high μ, and are now almost universally employed where a voltage amplifier at audio or carrier frequencies is required.

Miller effect

The effect of grid-anode capacity on the grid-cathode impedance is known as " Miller effect ". It can be shown that, with a resistive load, the input impedance is equivalent to a capacity

$$C = C_{gc} + C_{ga}(1 + M)$$

where C_{gc} and C_{ga} are the grid-cathode and grid-anode capacities respectively, and M is the gain of the stage.

If the anode load is reactive and a phase-shift θ occurs in the amplifier, the input impedance is equivalent to a condenser C in parallel with a resistance R, where :—

$$C = C_{gc} + C_{ga}(1 + M \cos \theta)$$

and
$$R = \frac{-1}{\omega C_{ga} . M \sin \theta}.$$

This resistive term will be negative if θ is positive, *i.e.*, for inductive loads, and may cause oscillations.

TETRODE VALVES

As was pointed out in a previous section, the screen-grid valve or tetrode was originally introduced to overcome ill-effects of grid anode capacity which become apparent when a triode is used as an RF amplifier. It has two grids between cathode and anode ; the grid nearer the cathode performs exactly the same function as the grid in the triode and is usually referred to as the " control grid ", while the additional grid acts as an electrostatic screen between the control grid and the anode, and is therefore called the " screen-grid " or " screen ". The screen is maintained at a high positive potential approaching that of the anode, and has a considerable effect on the electron stream between control grid and anode.

Consider the electrostatic field between the electrodes in terms of " lines of force ". If the screen were a solid metal plate maintained at a potential equal to that of the anode, lines of force leaving the cathode and grid would terminate on the screen, and there would be no electrostatic field in the space between screen and anode. Thus there would be no capacity between anode and screen, nor between grid and anode. Now consider the screen made up in the form of a close wire mesh and maintained at a potential not necessarily equal to, but approaching, that of the anode. This time the screening effect will be considerable but not perfect, though with a fine mesh structure it will be practically so. The result is that there will be capacity between the pairs of electrodes grid and screen, screen and anode, and anode and grid ; though the grid-anode capacity will be very much reduced from that in a triode. In commercial types of screen grid valve the residual grid-anode capacity varies from 0·001 μμF to 0·02 μμF, as compared with 2 μμF to 8 μμF for a triode.

Fig. 316 shows a typical anode characteristic for a tetrode, drawn under conditions of constant grid voltage (E_g) and constant screen voltage (E_{sg}). When the anode potential is zero, all the

emitted electrons are attracted to the screen, giving a fairly high screen current (I_{sg}) ; the anode current (I_a) will be zero. If, now, the anode potential is increased, some of the electrons passing through the meshes of the screen are carried on by their momentum and come under the influence of the anode, to which they are attracted, giving an anode current which will increase with

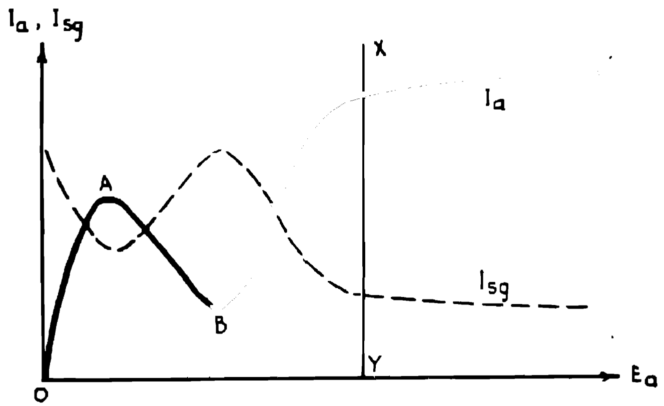


FIG. 316.—Anode characteristic of a screen grid valve or tetrode showing the effect of secondary emission.

increased anode potential. Due to the screening effect of the screen, however, the potential of the anode will have very little effect on the electrostatic field in the vicinity of the cathode, and an increase in anode potential will not appreciably increase the total space current. Any increase in anode current will therefore be at the expense of the screen current (*see* portion *OA* of curve shown in Fig. 316).

Secondary emission

As the anode potential increases, so also will the velocity of the electrons on arrival at the anode. The effect of bombarding the anode with fast moving electrons is that other electrons may be ejected by the force of impact. The quantity of electrons thus ejected (or " secondary electrons ", as they are usually called) will vary with the material of the anode and the velocity of the electrons reaching the anode from the cathode (" primary electrons "). In certain circumstances as many as ten secondary electrons may be liberated by one fast-moving primary electron. This phenomenon of " secondary emission " also occurs in the diode and triode, but in these cases secondary electrons emitted from the anode are attracted back into the anode and have no effect on the behaviour of the valve. With the tetrode, however, the velocity of the primary electrons is sufficiently high to cause secondary emission while the anode is at a lower potential than the screen grid ; there is therefore a tendency for the screen to collect these secondary electrons emitted

from the anode, thereby causing an increase in screen current at the expense of anode current. A further increase in anode potential will increase the velocity of the primary electrons and therefore increase the emission of secondary electrons ; if the screen is still at a higher potential than the anode it will collect practically all these slow-moving secondary electrons, with the result that the anode current will actually *decrease* with increased anode potential.

This state of affairs is represented by the portion AB of the anode characteristic (Fig. 316). Under the operating conditions that give this portion of the characteristic, the valve behaves as a *negative resistance* device, since an increase in anode voltage causes a decrease in the anode current, and a decrease in anode voltage an increase in the anode current. This negative resistance property is used in the dynatron oscillator (*see* Chapter 10).

If the anode potential is still further increased, the majority of the secondary electrons will no longer be attracted to the screen, but more and more will be drawn back into the anode and the anode current will once more increase with increased anode potential, at the expense of a decreasing screen current. The portion of the tetrode characteristic that is useful for most purposes is that portion well to the right of the vertical line XY in Fig. 316. In this region the curve becomes practically straight, and the anode current is nearly independent of anode voltage indicating a very high value of AC resistance R_a. The effect of the grid, however, is practically the same as if the screen and anode together formed the collecting electrode—that is to say, the mutual conductance is of the same

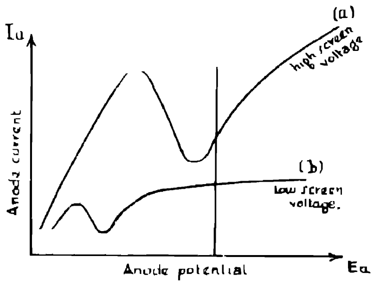


FIG. 317.—Anode characteristics for a tetrode showing effect of varying screen potential.

order as for a triode. Hence, from the relationship between the valve constants, the amplification factor μ is high. The mutual characteristics have much the same shape as those of a triode, but the curves for different values of anode voltage are closer together, indicating the high AC resistance.

The screen potential, as might be expected, has a considerable effect on the shape of the anode characteristic since it determines at what anode potential the effects of secondary emission will cease to be apparent ; it also determines the total space current, and hence the maximum standing anode current.

Fig. 317 shows anode characteristics for a typical tetrode plotted for high (curve *a*) and low (curve *b*) screen voltages. The screen voltage is obtained either by a tapping on the HT battery, or more commonly by connecting the screen to HT positive by means of a dropping resistor of suitable value or a potentiometer, as shown

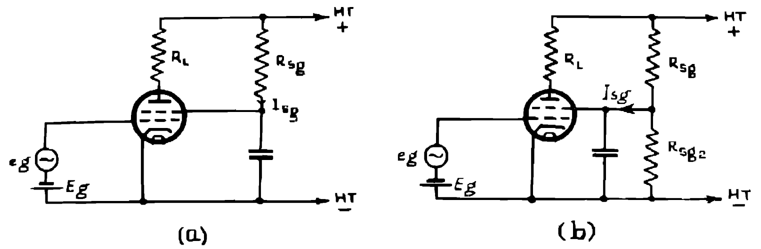


FIG. 318.—Simple circuit of a tetrode arranged as an amplifier.

in Fig. 318*a* and *b* respectively; the value of the resistance may be calculated from the fact that the screen current is normally about a quarter of the value of the anode current.

The screen-grid valve as an amplifier

Fig. 318 shows a tetrode arranged as a voltage amplifier. With an alternating voltage applied between cathode and grid, there will be fluctuations in screen current, just as there are fluctuations in anode current. The screen current will therefore cause fluctuations in the screen voltage due to the potential drop across R_{sg}. Since there is capacity between grid and screen of about the same order as between grid and anode for a triode there is still a leakage path for the transfer of energy from anode circuit to grid, thereby nullifying to a large extent the advantage of a low residual grid anode capacity. This effect is overcome by connecting the screen grid to the cathode through a condenser. This will represent a negligible impedance at high frequencies, and the screen grid and cathode will be virtually at the same (alternating) potential. There will then be no coupling between anode and grid circuits, apart from the very small residual grid-anode capacity. It should be noted that the advantages of low grid-anode capacity will be nullified unless great care is taken to screen the grid and anode circuits external to the valve. For this reason most valves are now metallised, and either the grid or the anode lead is brought out to the top cap to which a connection is made by means of a screened lead.

Due to the restriction on the working part of the characteristic imposed by secondary emission, the screen grid tetrode is of little or no use as a power amplifier, and its use as a voltage amplifier is limited, since it can handle only a very small grid swing. These valves are practically obsolete.

The critical-distance tetrode

In practice, in order to retain the essential features of the

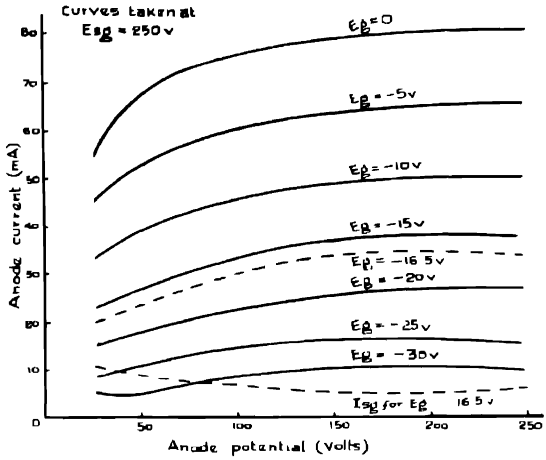


FIG. 319.—Anode characteristics of a critical-distance tetrode.

tetrode (*i.e.*, low grid-anode capacity and high amplification factor), and at the same time to increase its voltage and power handling capacity with limited anode voltages, it is necessary to suppress or reduce considerably the effects of secondary emission from the anode. There are several ways of achieving this

One way is to design the valve with a small grid-screen and large anode-screen separation combined with the use of an open-meshed screen The screen potential will still largely govern the space current, and the open-meshed screen will allow most of the electrons to pass through and come under the influence of the anode.

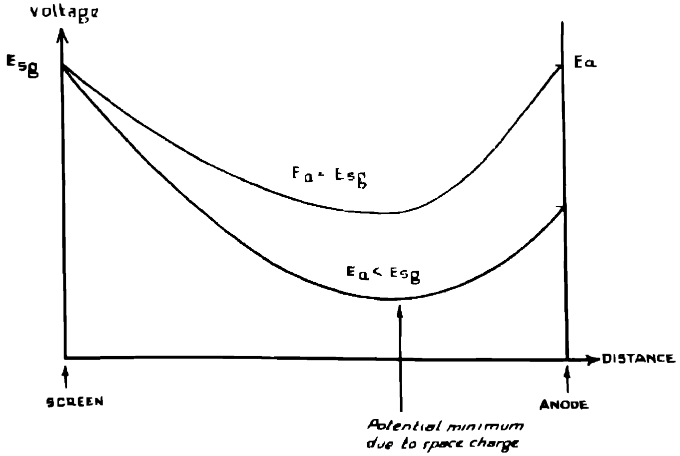


FIG 320.—Potential distribution between screen and anode in critical-distance or beam tetrode.

Even if the anode potential is very low compared with that of the screen, secondary electrons emitted at the anode will have to travel a considerable distance before the attraction of the screen is greater than that of the anode. Thus the effect of secondary screen current is entirely eliminated, and the anode characteristics of a "critical-distance tetrode" are as shown in Fig. 319. The screen current is made small by optical alignment of the wires of the control grid and of the screen, so that the latter will intercept a minimum number of electrons.

Fig. 320 shows the potential distribution between screen and anode for a critical-distance or beam tetrode.

The beam tetrode

To reduce further the effects of secondary emission additional electrodes may be provided between screen grid and anode.

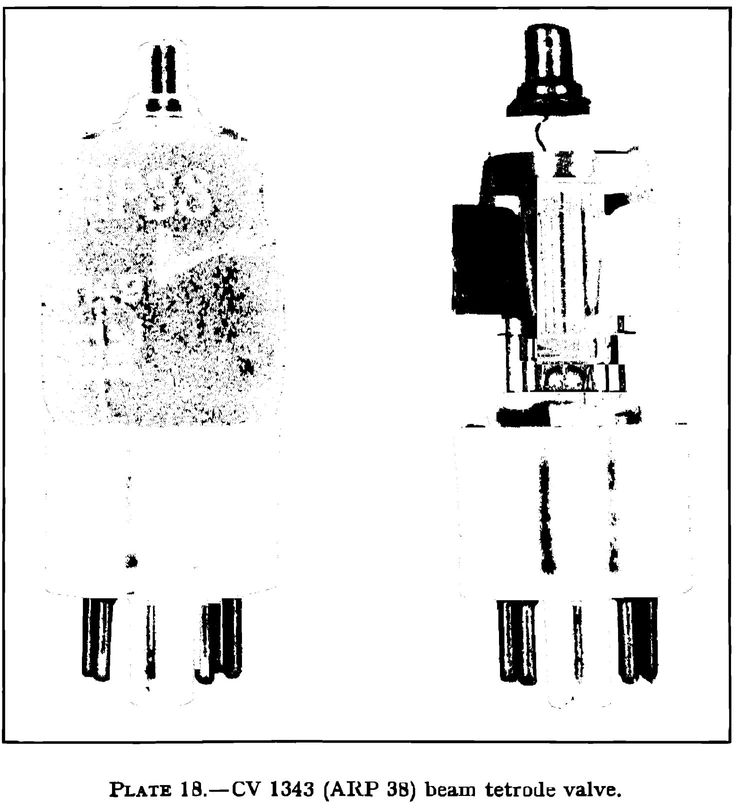


PLATE 18.—CV 1343 (ARP 38) beam tetrode valve.

These electrodes are connected to the cathode inside the valve and will repel the electron stream. The electrodes are arranged so that they concentrate the electron stream into a comparatively narrow beam (*see* Fig. 321), and for this reason are usually referred to as "beam-forming plates". The concentration of the electrons

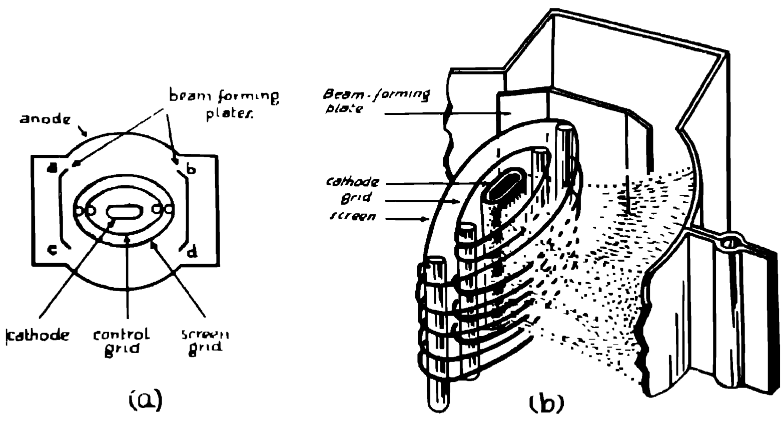


FIG. 321.—Arrangement of electrodes in the beam tetrode.

into a beam, combined with a large distance between screen and plate, gives an intensified space-charge effect in the screen-anode space which will repel the secondary electrons back into the anode. The screen current is again made small by having an open-meshed screen, and optical alignment of control grid and screen. Such a valve is called a "beam tetrode"—the 6V6 being a typical example. A set of anode characteristics for this valve is given in Fig. 322.

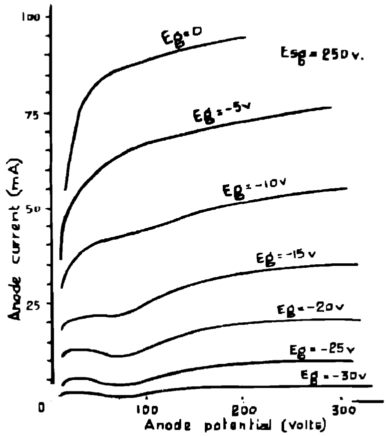


FIG. 322.—Anode characteristics of the 6V6.

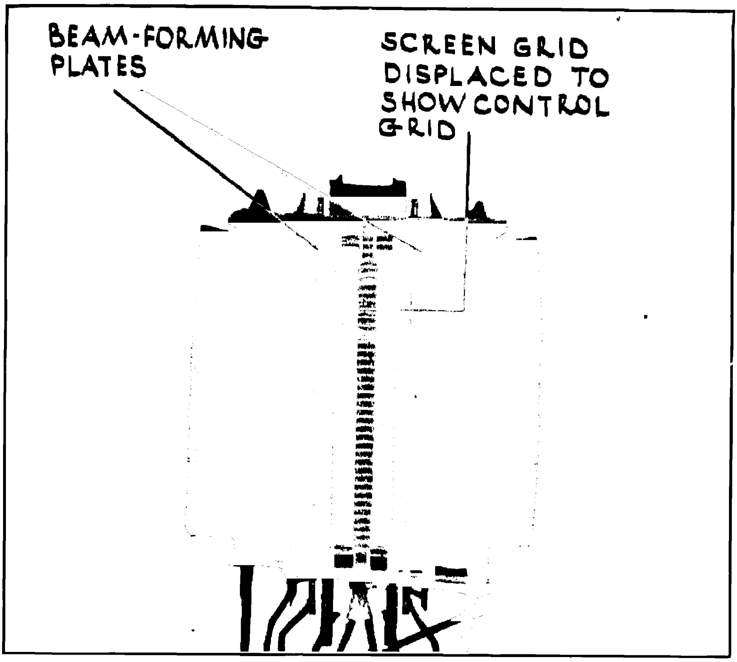


PLATE 19.— 6V6 beam tetrode valve, showing electrode structure.

Valves of the critical distance and beam tetrode types are often referred to as "kinkless" tetrodes for obvious reasons; in view of their suitability as power amplifiers they are also called "output tetrodes".

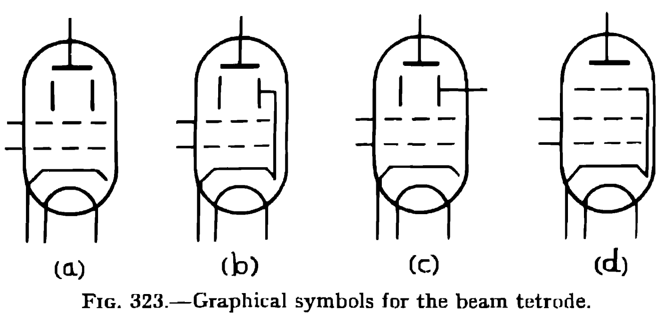


FIG. 323.—Graphical symbols for the beam tetrode.

In Fig. 323, *a*, *b* and *c* show the graphical symbols used for a beam tetrode. Fig. 323*a* is the general symbol, *b* is that used when internal connection is made between the beam-forming plates

and cathode, while *c* is used when the beam-forming plates are brought out to a separate terminal. The symbol of Fig. 323*d* is sometimes used, though strictly it refers only to a true pentode (*see* next section) that has the connection between the fourth electrode (suppressor grid) and cathode made internally.

PENTODE VALVES

An alternative method of reducing secondary emission is the introduction of an additional electrode, in the form of a third grid, between the screen and the anode. This third grid is called the "suppressor", and the resulting five-electrode or "pentode" valve is one of the most important types. The suppressor is given a negative potential relative to anode and screen, and this prevents the low-velocity secondary electrons from reaching the screen; it is usually built of very open-mesh wire so that it does not interfere appreciably with the passage of the high-velocity primary electrons towards the anode. The suppressor-grid is usually connected to cathode, but, since some other connection may require to be made (as, for example, in suppressor-grid modulation), the lead to the suppressor-grid is usually brought out of the valve to a separate pin and the connection made externally. In certain cases where a pentode is suitable only as a power amplifier the connection is made internally, and the pentode then has the same external connections as a tetrode (*see* Fig. 323*d*).

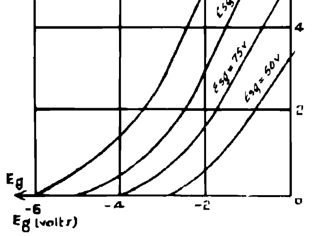


FIG. 324.—Mutual characteristics of a typical pentode (CV 1056)

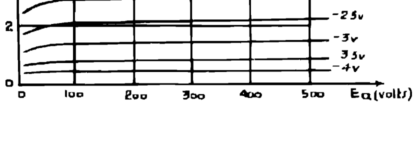


FIG. 325.—Anode characteristics of a typical pentode (CV 1056).

Fig. 324 shows the mutual characteristics, and Fig. 325 the anode characteristics, of a typical pentode, the CV 1056 (VR56) (EF36). The constants of such a valve are approximately :—

$$g = 2\cdot 4 \text{ mA/volt}\ ;\ R_a = 1\cdot 5 \text{ M}\Omega, \text{ and } \mu = 3600.$$

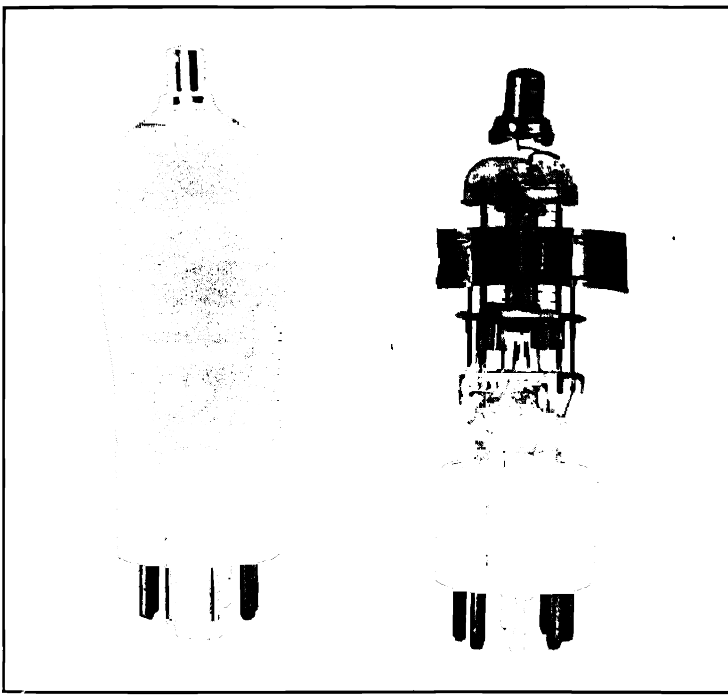


PLATE 20.—CV 1053 (ARP 34) (EF 39) pentode valve.

Variable-mu pentodes

The variable-mu pentode is a pentode in which the control grid is made to have an asymmetrical structure. This is normally done by making the pitch of the grid vary along its length, the meshes of the grid being closer at one end than at the other. An alternative method is to allow the cathode to project some distance outside the control grid. In either case the result is that various parts of the valve reach cut-off with different grid bias voltages, so that the overall cut-off comes gradually rather than abruptly.

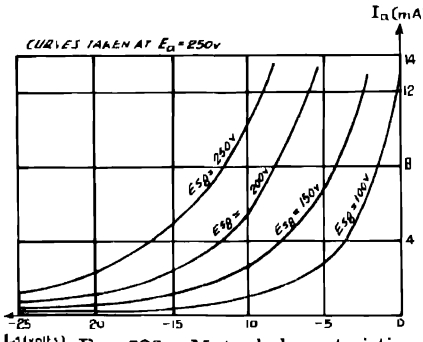


FIG. 326.—Mutual characteristics (for various E_{sg}) of variable-mu pentode (CV 1053).

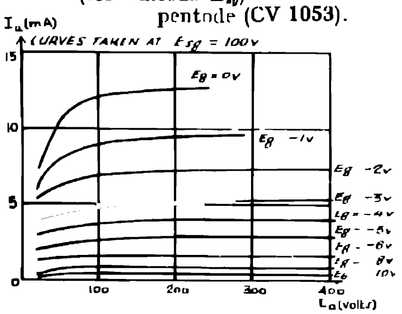


FIG. 328.—Anode characteristics (for various E_g) of variable-mu pentode (CV 1053).

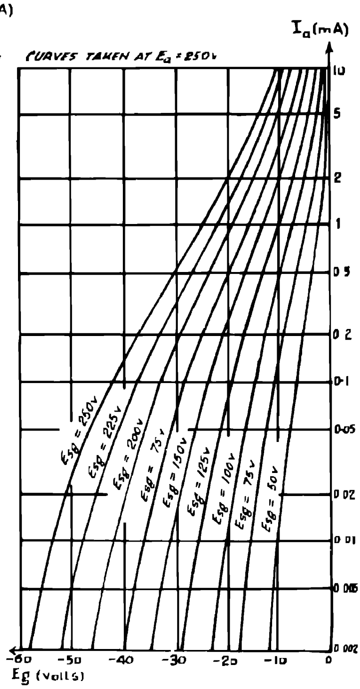


FIG. 327.—Mutual characteristics (for various E_{sg}) to log scale.

Fig. 326 shows a set of mutual characteristics of a typical variable-mu pentode CV 1053 (ARP34) (EF39), plotted for constant anode voltage, but variable screen voltage. As given in valve data sheets, such curves are often plotted with anode current on a logarithmic scale of microamps, in order to include the whole range of anode currents, and at the same time give an accurate picture of the lower part of the characteristic. A set of curves plotted to such a scale is given in Fig. 327 and does, of course, look entirely different from Fig. 326, but is nevertheless equivalent to it. Fig. 328 shows the anode characteristic of the same valve, plotted

for constant screen voltage but variable grid voltage. Fig. 329 shows mutual characteristics, plotted for constant screen and variable anode voltage; and Fig. 330 shows anode characteristics, plotted for constant grid and variable screen voltage.

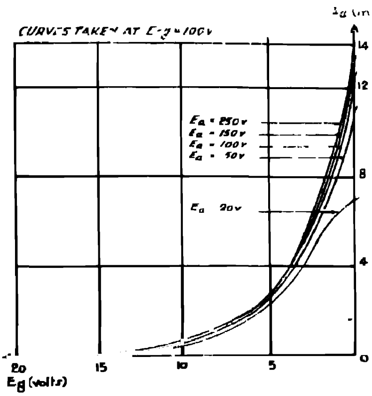


FIG. 329.—Mutual characteristics (for various E_a) of variable-mu pentode (CV 1053).

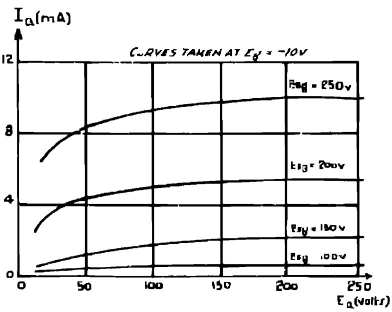


FIG. 330.—Anode characteristics (for various E_{sg}) of variable-mu pentode (CV 1053).

The object of a variable-mu valve is that it enables the stage gain of a voltage amplifier stage to be varied over a wide range by varying the bias voltage on the control grid, and therefore the "g" of the valve.

The pentode as an amplifier

The general principles of amplifier stages using either tetrodes or pentodes are the same, but they differ from those employed for

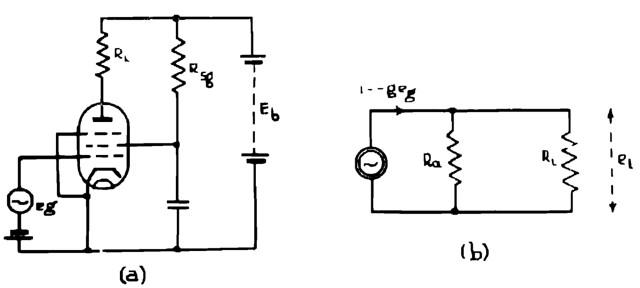


FIG. 331.—Pentode arranged as a voltage amplifier, with constant current generator form of equivalent circuit.

the triode amplifier stage. For the moment all reference to power amplifiers will be omitted, and attention concentrated on voltage amplifier stages.

Fig. 331a shows a pentode arranged as a voltage amplifier, and Fig. 331b shows the constant-current generator form of the

equivalent circuit. When using a triode as a voltage amplifier a load resistance somewhat higher than the AC resistance of the triode is chosen in order to get as high a value of stage gain as is reasonably possible consistent with the available HT supply voltage and the drop across the anode load. With the pentode, the value of R_L is limited by supply voltage considerations to a value that is normally small compared with the AC resistance of the valve. For example, a reasonable value of anode load for a pentode having an AC resistance of 1·5 MΩ may be of the order of 100 kΩ.

For pentodes, the constant-current generator form of equivalent circuit (Fig. 331*b*) is therefore preferable to the constant-voltage generator form.

Clearly the stage gain is given by :—

$$M = \frac{e_L}{e_g} = g\frac{R_a R_L}{R_a + R_L} \qquad \text{(see equation 20)}$$

i.e.

$$M = g\frac{R_L}{1 + \dfrac{R_L}{R_a}} \qquad (21)$$

If, as is usually the case, R_a is very much larger than R_L, the result for a pentode may be written as :—

$$M = g.\ R_L \qquad (22)$$

Now suppose that the pentode of Fig. 331*a* is a CV 1091 (ARP35) (EF50), anode characteristics of which are given in Fig. 332.

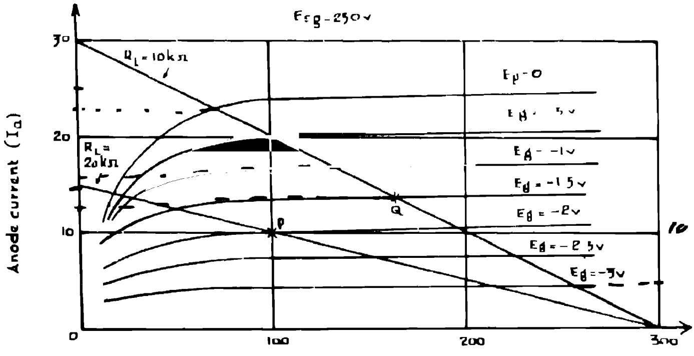


FIG. 332.—Anode characteristics of a pentode (CV 1091).

Suppose that an HT voltage supply of 300 volts is available, and that it is decided to try a load resistance of 10 kΩ. Using the same method as for a triode (p. 341), one can draw the load line, and choose the operating point Q. From the equal intercepts it can be seen that grid swings up to 1·5 volts may be handled without undue distortion.

The peak value of the alternating component of the anode current corresponding to the 1·5 volt signal will be 9·5 mA, giving

a peak voltage of 95 volts across the anode load. Thus the stage gain (M) would be :—

$$M = \frac{95}{1 \cdot 5} = 63$$

Using the formula of equation 22, this result can be obtained easily, for from the valve data $g = 6 \cdot 5$ mA/volt.

i.e. $\quad g = 6500$ µmhos

$\therefore \quad M = gR_L = 6500 \times 10^{-6} \times 10^4 = 65$

which is approximately the same result as before.

If an anode load of 20 kΩ is chosen, another load line is obtained with the operating point P This time the intercepts indicate that

PLATE 21 —CV 1091 (ARP 35) (EF 50) pentode valve.

a peak signal of only 0 5 volts may be applied without undue distortion. The peak value of alternating anode current corresponding to the $0 \cdot 5$ volt peak signal is $3 \cdot 25$ mA, and the peak alternating voltage across the anode load is 65 volts. Thus the stage gain is given by :—

$$M = \frac{65}{0 \cdot 5} = 130$$

Alternatively, using equation 22, since $g = 6500$ µmhos and $R_L = 20$ kΩ :—

$$M = 6500 \times 10^{-6} \times 20 \times 10^3 = 130$$

which agrees with the figure previously obtained.

The greater the anode load the greater the voltage gain, just as in a triode, but an increase in anode load may give a considerable increase in distortion, unless the signal applied to the grid is very small. This question of distortion will be considered in more detail when considering power amplifiers. For voltage amplifiers, the input signal voltages are usually very small, and the highest possible load is taken consistent with the available HT voltage and the magnitude of the signal to be amplified.

MISCELLANEOUS VALVE TYPES

Frequency-changer valves

If the potential on the suppressor grid of a pentode is varied, there will be a corresponding variation in the anode current. This

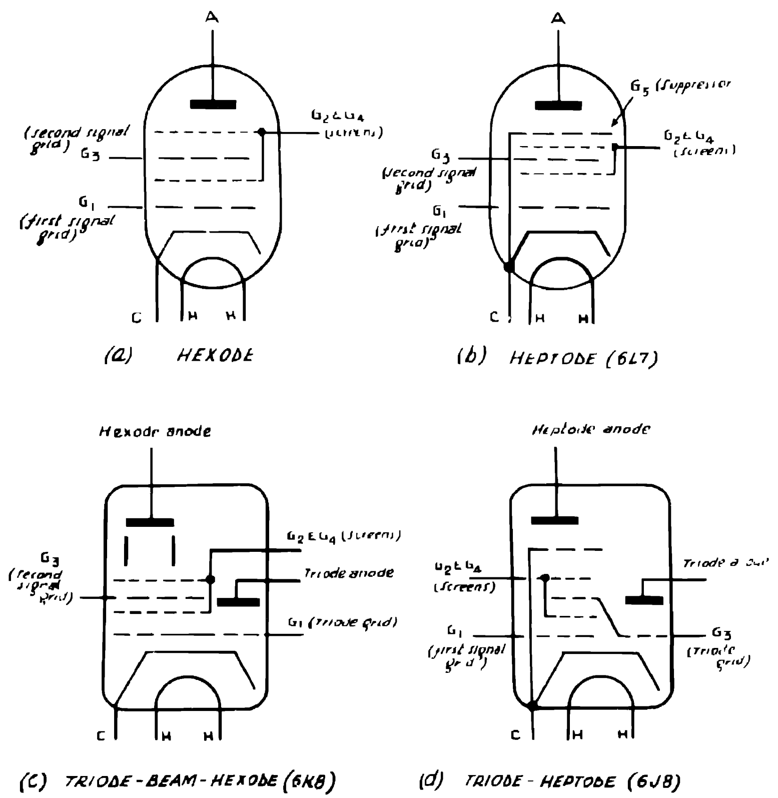


FIG. 333.—Frequency-changer valves.

fact can be used to allow two independent signal voltages simultaneously to control a single anode current, by applying one voltage to the control grid and the other to the suppressor. Owing to the

non-linearity of the valve, intermodulation between the frequencies of the two inputs will take place, as will be seen in Chapter 11. The anode-suppressor capacity of a pentode used in this way has the same ill-effects as the anode-grid capacity of a triode; special "frequency-changer" valves, as they are called, have therefore been developed with electrostatic screening between the anode and the second signal grid.

The hexode

The hexode is a modified pentode valve, having an extra electrostatic screen interposed between the third ("suppressor", or "second signal-") grid and the anode (*see* Fig. 333*a*). This additional screen (G_4) is usually connected internally to the existing screen (G_2), and is maintained at a steady positive potential, so that it functions in exactly the same manner as the screen of a tetrode, and the hexode forms a much more stable frequency-changer than the pentode.

The heptode

Since, in the hexode, a screen at positive potential is adjacent to the anode, trouble may be experienced from secondary-emission effects as in the case of the tetrode (*see* p. 351). This difficulty can be overcome either by employing the principle of the beam tetrode, or by the insertion of a suppressor grid between screen and anode. In the first case, the resulting valve is called a "beam hexode", and in the latter, a "heptode" (*see* Fig. 333*b*, which represents a type 6L7).

Triode hexodes and triode-heptodes

Frequency-changer valves are normally used to intermodulate two signals, one of which is generated by an oscillator coupled directly to one of its two control or signal grids. To reduce the number of valves needed in a piece of equipment, hexodes and heptodes are sometimes built into an envelope that also contains a triode; this triode can be used as an oscillator, and has its grid internally connected to one of the two signal grids of the hexode or heptode. Fig. 333*c* shows a representation of a triode-beam-hexode, such as the type 6K8, in which the triode grid is internally connected to the hexode grid nearest the cathode, while Fig. 333*d* shows a triode-heptode (*e.g.* the 6J8), in which the triode grid is connected to the third grid (G_3).

The co-planar-grid valve

The co-planar-grid valve, exemplified by the 4045A (*see* Fig. 334), is sometimes used in the output stage of multi-channel carrier telephone equipment.

It is virtually a tetrode in which the screen grid and the control grid are intermeshed so that they are co-planar and equidistant from the cathode. In operation, the control grid (*a*) is given a negative bias of 60 volts so that it can handle a very large grid

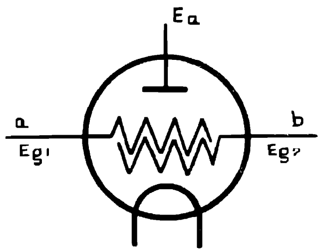


FIG 334 — Co-planar valve (4045 A)

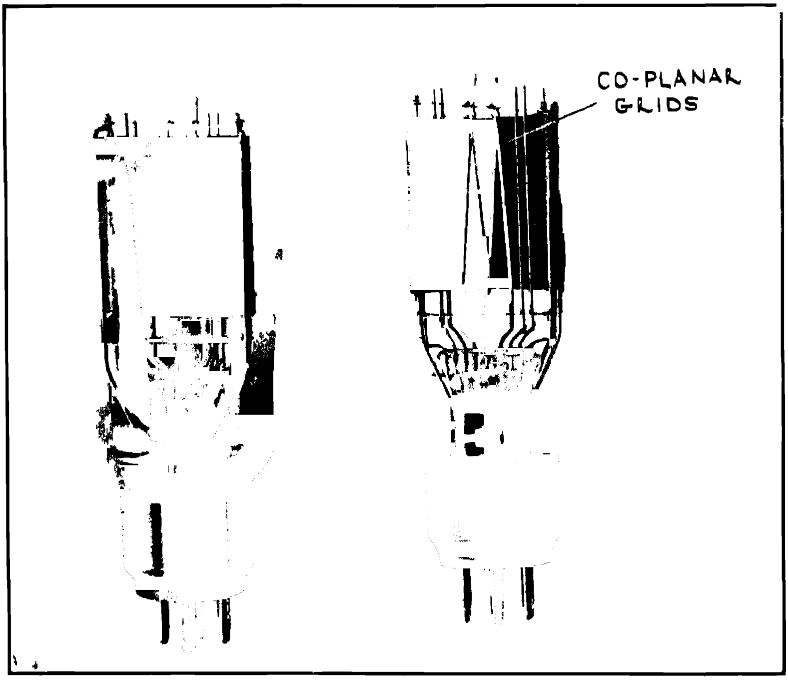


PLATE 22 —4045A Co planar grid valve.

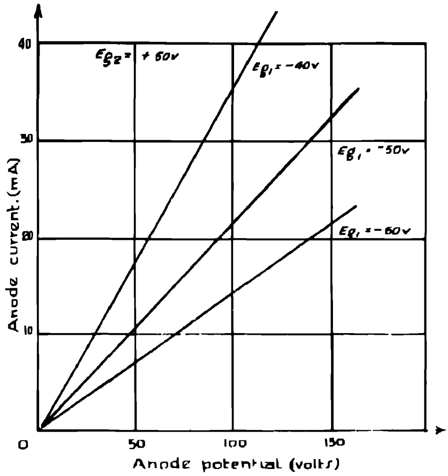


FIG 335.— Anode characteristics of the Co-planar valve (4045 A).

swing without becoming positive; the co-planar-grid (b) is given a positive bias of about 60 volts. As far as anode current is concerned, the valve behaves as a triode in which the grid potential is zero—that is, the anode current is large. Fig. 335 shows anode characteristics for this valve.

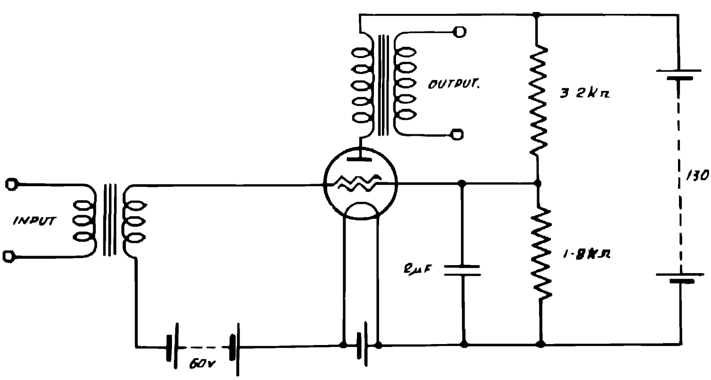


FIG. 336. –Co-planar valve employing grid bias battery.

METHODS OF BIASING

Bias

It has been seen that a steady negative bias must be applied to the grid of a valve used as an amplifier, in addition to the signal voltage. The various ways in which this can be done will be considered in the ensuing paragraphs. It will be noted that, in every case, a DC path exists between grid and cathode; this is

most important, for as has been seen (page 348), if an open-circuit or even a high resistance path exists between grid and cathode, damage may result.

Battery bias

So far it has been assumed that the grid bias voltage is obtained from a battery in the grid circuit. This is known as " battery bias ", and is only one of the various methods by which a steady potential difference may be maintained between cathode and grid. The method is seldom used nowadays, about the only example in line equipment being the — 60 volts grid bias for the co-planar valve (*see* Fig. 336).

Filament current bias

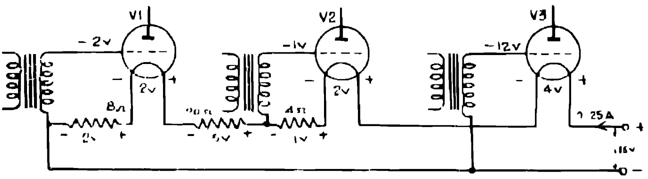


FIG. 337.- Method of obtaining filament current bias.

This is a convenient method for obtaining bias with directly-heated valves, particularly when a number of valves are operated with their filaments in series, an example being shown in Fig. 337.

The three valve filaments are in series with various resistances, and the whole filament circuit draws 0·25 amps filament current. Valves V1 and V2 take 0·25 amps at 2 volts, and V3 takes 0·25 amps at 4 volts. The voltage drops across the various filaments and resistances are shown in the diagram. It is easily seen that the potentials of the various grids relative to the negative ends of the respective filaments are as shown in the diagram, and that these are maintained by the flow of filament current.

Bias from HT supply

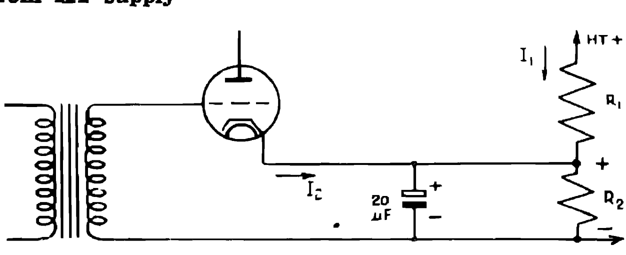


FIG. 338.—Method of obtaining bias from HT supply.

If a large grid bias voltage is required, the cathode of an indirectly-heated valve may be made positive with respect to the grid by employing a potentiometer across the HT supply, as shown in Fig. 338. When determining the values of R_1 and R_2, it should

be remembered that R_2 carries not only the potentiometer current I_1 but also the cathode current I_2 of the valve. The bias applied is therefore $(I_1 + I_2)\ R_2$.

The disadvantage of this method is that the resultant HT voltage available to be applied between anode and cathode of the valve has been reduced by the amount of grid bias voltage applied.

Cathode bias

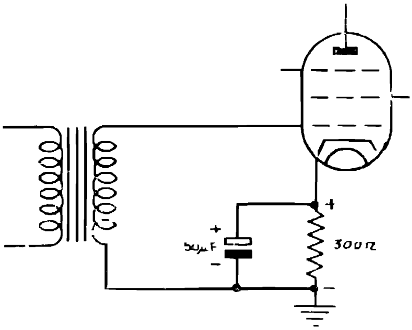


FIG 339 —Method of obtaining cathode current bias.

Fig. 339 shows the method of producing bias that is most often used with indirectly-heated valves. A resistance—in this case 300Ω—is inserted in the cathode lead of the valve; this resistor has to carry the whole of the standing cathode current, *i.e.* the sum of the steady anode and screen currents. Across this resistor therefore will be developed a steady voltage in the sense shown; this will

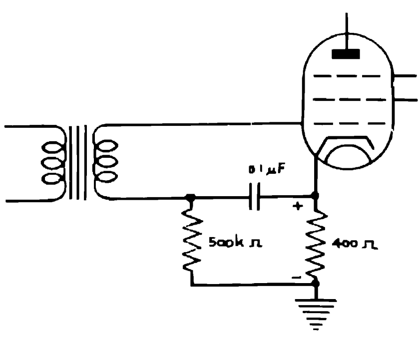


FIG. 340 —Another method of decoupling the cathode resistor.

make the grid negative with respect to cathode, the amount of bias depending on the value of cathode resistance and on the cathode current. The large capacity condenser shunting this resistor is provided for decoupling—that is, providing an alternative low impedance path for the oscillatory currents. If the condenser were

not connected, the variations in anode current would cause variations in the potential across the cathode resistor; these would be fed on to the grid and would be of such a phase as to oppose the applied signal. The omission of the cathode decoupling condenser is a convenient way of obtaining "current negative feedback", and is fully discussed in Chapter 9. If bias only is required, the cathode resistor must be decoupled, otherwise both bias and current negative feedback are obtained.

Another example of the provision of cathode bias without current negative feedback is shown in Fig. 340.

The only essential difference between this circuit and the last is in the method of decoupling. A large resistance is included in the grid circuit, and this enables the cathode resistor to be efficiently decoupled using quite a small decoupling condenser. At 1600 c/s, the reactance of the $0 \cdot 1$ μF condenser is in the region of 1000Ω. Thus the voltage developed across this condenser and applied between cathode and grid will be only $\frac{1000}{\sqrt{500{,}000^2 + 1000^2}} \simeq \frac{1}{500}$th of the alternating voltage developed across the cathode resistor.

Grid leak bias

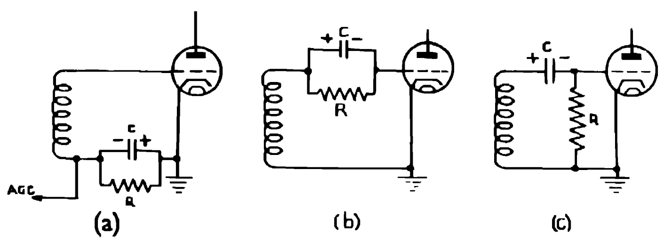


FIG. 341.—Grid leak bias.

This form of biasing, which depends on the flow of grid current for the production of the bias voltage, is frequently employed in oscillatory circuits. Fig. 341 shows three circuits for producing this type of bias; all three are in common use and the action is the same in each case. Consider Fig. 341*a*; when a signal is applied to the input, the grid will become positive with respect to the cathode on every positive half-cycle. Grid current will flow, developing a voltage across the resistor *R* and charging up the condenser *C* in such a direction as to bias the grid negatively. During the negative half-cycles, no grid current will flow, and the condenser *C* will start to discharge through the resistor *R*. If the time constant of *C* and *R* is large compared with the periodic time of the input signal, the condenser will retain its charge from one positive half-cycle to the next, and a steady negative bias will be produced. Equilibrium will be set up with the valve running into grid current on the peaks of the positive half-cycles (*see* Fig. 342), the grid current flowing being just sufficient to maintain the charge on the condenser.

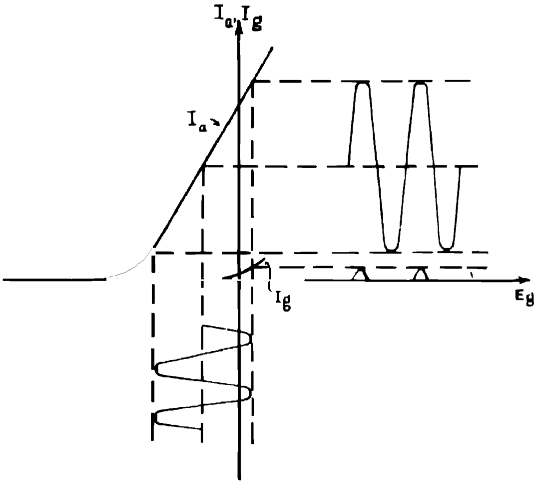


FIG. 342.—Explaining operation of grid leak bias.

This circuit is equivalent to that shown in Fig. 343*a*; for the valve will pass current from grid to cathode, but not in the reverse direction, and it can thus be regarded as a rectifier. This is a half-wave rectifier circuit, and it produces a DC voltage across C as shown in Fig. 343*b*. If the ripple in the output is to be small, it follows that the time constant CR must be greater than the period of one cycle. If CR is large, the voltage across C is almost constant; and as has been seen, this voltage constitutes the bias on the grid of the valve.

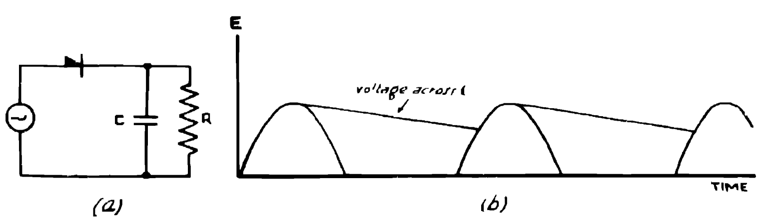


FIG. 343.— (*a*) Equivalent circuit for grid leak bias, and (*b*) Voltage across grid condenser.

If the amplitude of the signal on the grid increases, the bias will also increase to the new peak value; and if the signal decreases, the bias will decrease, though the rate at which the bias follows the signal will depend on the time constant CR. If this method of biasing is used with a variable-mu value, a type of Automatic Gain Control (AGC) results, for the larger the input, the larger the bias, and the smaller the amplification factor of the valve and the stage gain. In this way the amplitude of the output can be kept more

or less constant. In the arrangement of Fig. 341*a* the voltage developed across *C* and *R* can be applied as AGC bias to other stages.

DECOUPLING

The circuits of Fig. 344 and Fig. 345 have already been discussed, but they are now repeated for emphasis. As regards Fig. 344 it was stated (p. 353) that fluctuations in screen current would cause the potential of the screen to vary with respect to cathode, and would eventually feed back into the grid circuit unless a condenser *C* is

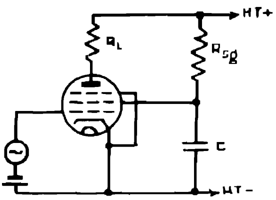


FIG 344.—Decoupling of screen circuit in a pentode or tetrode.

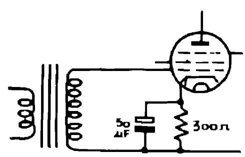


FIG. 345 —Decoupling of cathode resistor.

inserted between cathode and screen to provide a low impedance path for the alternating screen current and prevent it flowing through R_{sg}. The condenser *C* is said to *decouple* R_{sg}. In Fig. 345 the cathode resistor is common to both grid and anode circuits, and it was noted that if the alternating component of the anode current were allowed to flow in it, the resulting alternating voltage would be fed into the grid circuit. It was also noted that if the cathode resistor were shunted by means of a large capacity condenser, the alternating component of the anode current would have an alternative low impedance path, and very little alternating voltage would be applied to the grid circuit. The 50 μF condenser is said to *decouple* the cathode resistor.

Decoupling, then, is effected by providing an alternative low impedance path for the alternating components in order to prevent alternating voltages being developed across an impedance. An impedance that is common to two or more circuits will always introduce interaction between those circuits, unless it is decoupled.

Grid circuit of directly heated valve

Fig. 346*a* shows the circuit of an amplifier employing filament current bias. This filament current bias is derived from the 4-volt potential drop across the 16Ω resistor in the filament circuit. This particular filament circuit is connected across the LT supply in parallel with a vibrator, which produces disturbing voltages across the LT supply.

In a directly-heated valve the potential of the cathode varies along its length, so that if the length of the filament is divided into

a large number of short elements, each of which may be regarded as equipotential throughout its length, then the potential of the anode relative to the "cathode" will be greater for an element near the negative end of the filament than for one near the positive end. Similarly the grid will be more negative with respect to the positive end of the filament. The result of these two factors is that the various elements of the filament contribute unequally to the anode current (*see* Fig. 347), $E_g + \frac{E_a}{\mu}$ being greater for elements near the negative end of the filament. The total anode current may be regarded as a function of $E_g' + \frac{E_a'}{\mu}$, where E_g' and E_a' are the grid and anode potentials respectively relative to a point, called the "mean emission point", which is off-set towards the negative end of the filament. The position of the mean emission point varies for different valves, but is generally about one-third of the way along the filament from the negative end.

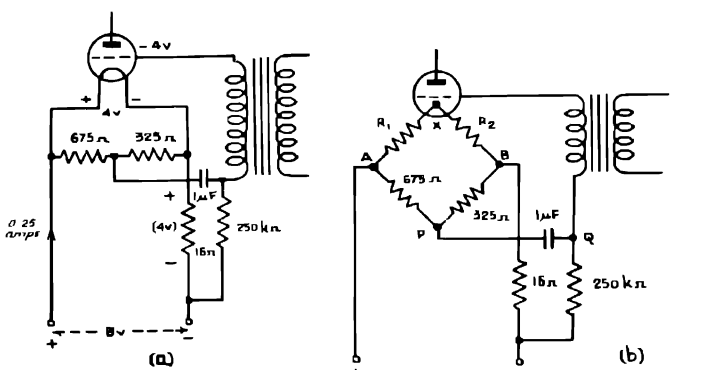


FIG. 346.—(*a*) Decoupling of filament circuit in the case of a directly-heated valve.
(*b*) Equivalent circuit showing the mean emission point.

Returning now to Fig. 346*b*. The point X is the mean emission point of the filament. All emission will be assumed to take place at this point. The resistors AX and XB represent the resistance of the filament on either side of this mean emission point, and the two resistors AP and PB are chosen so that $APBX$ is a balanced Wheatstone bridge,

$$i.e. \quad \frac{AX}{XB} = \frac{AP}{PB}$$

$$i.e. \quad \frac{R_1}{R_2} = \frac{675}{325}$$

Thus if the vibrator circuit applies an alternating voltage across AB, the point P will be at the same potential as the mean

emission point X. The 1 μF condenser connecting P and Q provides a low impedance path at the vibrator frequency and ensures that there is no alternating potential difference between Q and P, and therefore none between Q and X. Thus as far as the disturbing alternating voltages set up by the vibrator are concerned, the

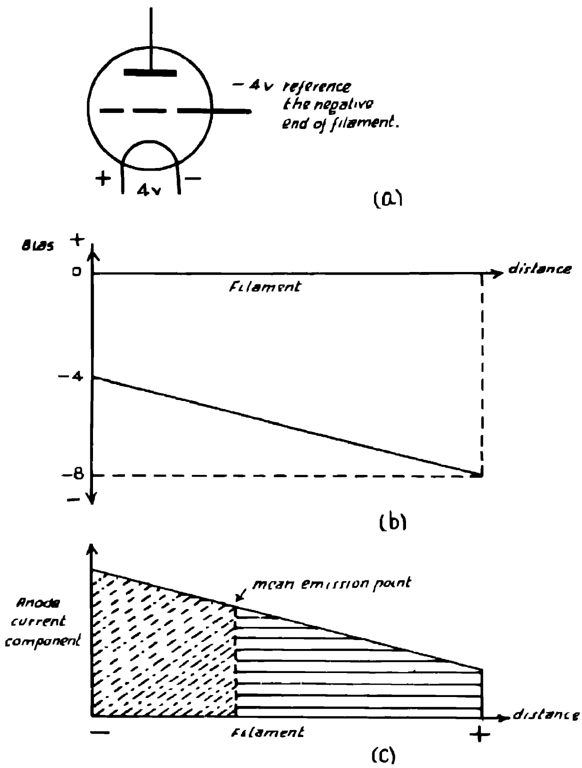


FIG. 347.— Illustrating mean emission point of directly-heated filament.

grid is at the same potential as the mean emission point of the filament; the space charge and hence the anode current will therefore be independent of these disturbing voltages. The grid circuit in this case is said to be decoupled to the mean emission point of the filament.

Anode circuit decoupling

Fig. 348 shows the arrangement of an amplifier stage that is one of several sharing the same HT supply. In this case the internal resistance of the battery (or other form of supply) is an impedance common to two or more anode circuits (and also screen circuits, if pentodes are used); to prevent coupling between stages, it is necessary that the resistance of the battery be decoupled to prevent alternating components of the various anode currents flowing through it. A resistance R is placed in each anode circuit, and the

condenser C provides an alternative path whose impedance is low compared with R. In addition, since the HT supply to this stage is fed *via* the potentiometer formed by R and C, alternating components from other circuits connected to the same power supply will be reduced to negligible amplitudes, provided that R is large compared with the reactance of C at the frequencies concerned.

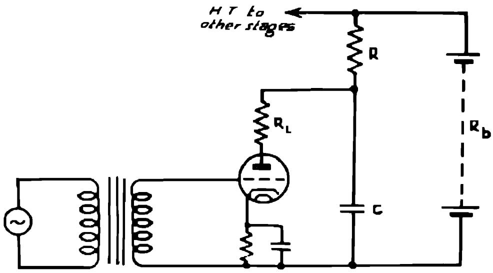


FIG. 348.—Decoupling of an HT supply common to several stages.

The values of C and R used may vary with the degree of decoupling required, which will depend on the number of stages sharing the common supply, and the relative importance of each stage as a disturbing influence on the others. Typical values are 5000Ω and 2 μF for audio frequencies; at carrier frequencies a smaller capacity condenser may provide a sufficiently low impedance alternative path, and 5000Ω and 0·5 μF are typical values. The degree of decoupling is increased by increasing both R and C—*e.g.* 20kΩ and 4 μF may be used in an audio frequency amplifier where decoupling is particularly important.

CHAPTER 8

VALVE AMPLIFIERS

Valve amplifiers may be divided into " voltage amplifiers " and " power amplifiers "—*i.e.* those which are designed to deliver a large *voltage* (but no current), as when feeding the grid of a subsequent stage ; and those which are designed to deliver actual *power*, as in the case of an output stage feeding a loudspeaker. Those within either of these classes can be sub-divided into " wide-band amplifiers " and " narrow-band amplifiers ".

Wide-band amplifiers are designed to give more or less uniform amplification over a wide frequency range, which may be from very low frequencies (of the order of 100 c/s), up to a maximum that may be anywhere from 3 kc/s to several Mc/s for line equipment amplifiers. For convenience in treatment, wide-band amplifiers may be further subdivided according to the type of coupling, of which resistance-capacity coupling is the most important.

Narrow-band amplifiers are designed to give amplification over a narrow band of frequencies, *e.g.* from 480 to 520 c/s, or from 460 to 470 kc/s—the word " narrow " meaning, in this connection, that the difference between the highest and lowest frequencies amplified is small compared with the mid-band frequency.

RESISTANCE-CAPACITY COUPLED VOLTAGE AMPLIFIERS

Fig. 349*a* shows a simple resistance-capacity coupled stage; when a signal is applied to the grid of V_1, alternating voltages are

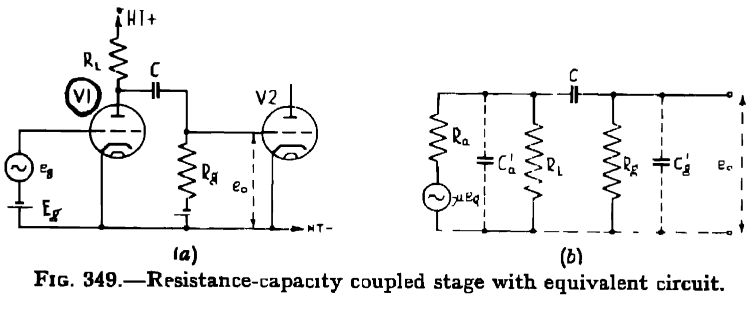


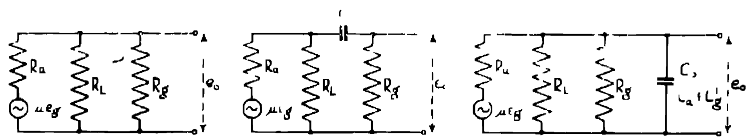
FIG. 349.—Resistance-capacity coupled stage with equivalent circuit.

developed across the load resistance R_L, which is effectively between anode and cathode of V_1. Anode and cathode are connected by the circuit consisting of C and R_g in series, and the voltage across R_g is applied between grid and cathode of V

Equivalent circuit in constant-voltage generator form

The equivalent circuit in the constant-voltage generator form is shown in Fig. 349*b*, and it will be noted that two capacities C_a' and C_g' have been included which do not appear in the original circuit diagram. C_a' is the output capacity of V_1, made up of anode-cathode capacity plus any stray wiring capacity to the left of the coupling condenser C, and C_g' is the input capacity of V_2 plus stray capacities to the right of the coupling condenser.

From the equivalent circuit it is clear that, since the network contains capacity, the voltage e_o applied to the second stage will vary with frequency, and therefore the stage gain $\frac{e_o}{e_g}$ will also vary with frequency. A complete analysis of the behaviour of this network is difficult, but the analysis can be simplified by dividing the range of frequencies into three parts.



(*a*) (*b*) (*c*)

FIG 350 —Equivalent circuits of R C coupled stage for medium, low and high frequencies.

Over a certain intermediate range of frequencies the small capacities C_a' and C_g' will have a high reactance and will not appreciably shunt the resistors R_L and R_g respectively. Also, the series condenser C will have only a small reactance, which can be neglected in comparison with R_g. The resulting equivalent circuit of Fig. 350*a* is therefore substantially correct for medium frequencies; since this is a purely resistive circuit, the stage gain will be independent of frequency over the range where the simplification is valid, and this range clearly depends on the values of R_L, R_g, C, C_a' and C_g'.

At lower frequencies, where this simplification is not admissible, the reactance of C_a' will still be negligible as a shunt on R_L, and C_g' negligible compared with R_g, but the reactance of C may be quite large and will increase with decrease in frequency. The equivalent circuit for low frequencies is shown in Fig. 350*b*. Since the reactance of C increases with decrease in frequency, and C and R_g form a voltage divider across R_L, e_o and hence the stage gain will decrease as the frequency decreases.

At high frequencies, the reactance of C will be negligible compared with R_g, but the capacities C_a' and C_g' will together form an appreciable shunt on R_L and R_g; this shunting capacity is denoted by C_s in Fig. 350*c*, which is the simplified equivalent circuit applicable for high frequencies. The shunting effect of C_s across the load becomes more marked, and causes a reduction in the stage gain at higher frequencies.

Variation of gain with frequency

Fig. 351 shows the way in which stage gain varies with frequency in a typical R-C coupled stage. The three ranges of frequency are clearly seen here, although the boundaries are of course not sharply

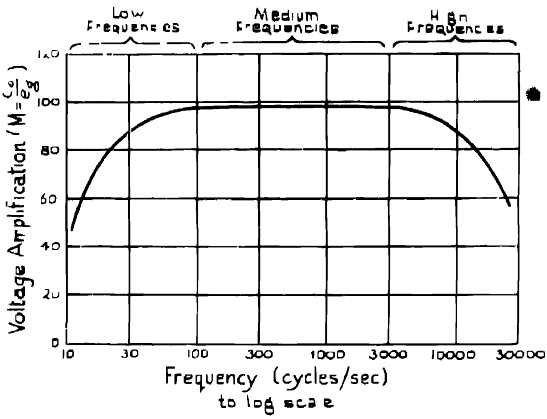


FIG. 351 —Gain-frequency response of a typical R-C coupled stage.

defined. In particular, the medium frequency range, over which the stage gain is substantially constant, is seen to be from about 100 c/s to 3000 c/s.

From Fig. 350*a*, the stage gain at medium frequencies is seen to be :—

$$M = \mu \cdot \frac{R_L'}{R_a + R_L'} \qquad (1)$$

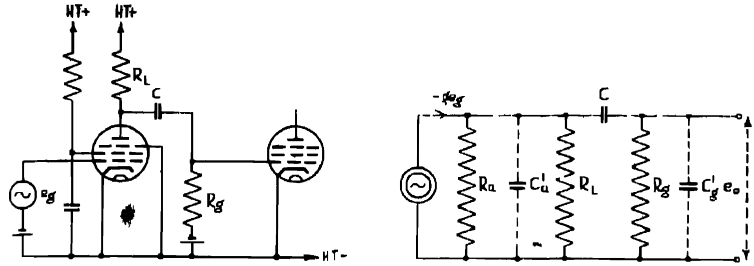
where
$$R_L' = \frac{R_L R_g}{R_L + R_g} \qquad (2)$$

Thus M is always less than μ, but may be increased, subject to this limitation, by an increase of R_L', and equation 2 shows that, for a high value of R_L', both R_L and R_g must be large.

Suppose, then, that an effort be made to secure a high stage gain by selecting very large values for R_L and R_g. At high frequencies the shunting effect of C_s (*see* Fig. 350*c*) will become appreciable at much lower frequencies than before ; that is to say, the stage gain will begin to fall off at much lower frequencies. Thus although the stage gain at middle frequencies has been increased, this increase is obtained at the expense of a reduction in the frequency range over which the response is flat. The design of an R-C coupled amplifier is therefore a compromise between the stage gain at middle frequencies, and the frequency range over which the response is flat ; one can be increased only at the expense of the other.

R-C stage using pentode valves

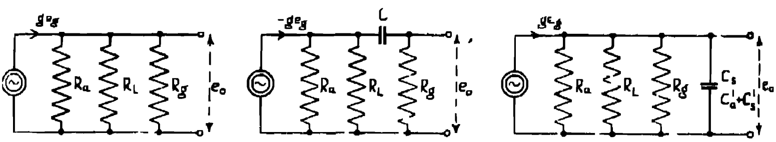
Fig. 352*a* shows a simple R-C coupled stage using pentodes,



(a) (b)

FIG 352 —Resistance-capacity coupled amplifier employing pentode valves.

and Fig. 352*b* the corresponding equivalent circuit in the constant-current generator form. Figs 353*a*, *b* and *c* show the simplified equivalent circuits for medium, low and high frequencies respectively.



(a) (b (c)

FIG 353 —Equivalent circuits of Fig 352*a* for medium, low and high frequencies

In the case of medium frequencies (Fig. 353*a*) :—

$$M = g \cdot R_{eq} \tag{3}$$

where
$$R_{eq} = \frac{1}{\dfrac{1}{R_a} + \dfrac{1}{R_L} + \dfrac{1}{R_g}}$$

For a pentode, R_a is large compared with R_L and R_g in parallel, so that only a small error is introduced by neglecting $\dfrac{1}{R_a}$ compared with $\left(\dfrac{1}{R_L} + \dfrac{1}{R_g}\right)$

$$\therefore \qquad R_{eq} \simeq \frac{R_L R_g}{R_L + R_g} \tag{4}$$

If, as is frequently the case, R_g is much larger than R_L (*e.g.*, $R_L = 30$ kΩ, $R_g = 0 \cdot 5$ MΩ) this can be simplified still further, giving :—

$$R_{eq} \simeq R_L$$

Thus for a pentode, with the proviso that $R_L \ll R_g$:—

$$M = g \cdot R_L \tag{5}$$

It will be clear from Fig. 353*b* and *c* that the stage gain decreases

at low and at high frequencies, and that the general shape of the gain-frequency response is the same for a pentode as for a triode (Fig. 351).

Stage gain of R-C stage at low and high frequencies

A rather more detailed examination will now be made of the stage gain of an R-C coupled stage at low and high frequencies.

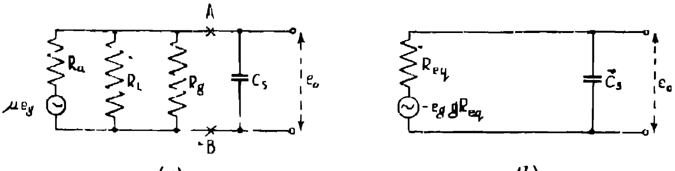


FIG 354 — Equivalent circuits of an R-C coupled stage at high frequencies.

Fig. 354*a* shows the equivalent circuit for high frequencies. By Thévenin's theorem, that part of the circuit to the left of AB can be replaced by a simple generator. The impedance of this generator is the impedance looking into AB to the left, *i.e.*, R_{eq} and the EMF of the generator will be the open circuit voltage at AB, *i.e.*, the voltage output of the stage over the medium frequency range, *i.e.*, $-e_g \cdot g \cdot R_{eq}$ (using equation 3). Thus Fig. 354*b* is another form of the equivalent circuit. From this circuit :—

$$e_o = -e_g \cdot g \cdot R_{eq} \cdot \frac{\frac{-j}{\omega C_s}}{R_{eq} - \frac{j}{\omega C_s}}$$

i.e.
$$\frac{e_o}{e_g} = -g \cdot R_{eq} \cdot \frac{1}{j\omega C_s R_{eq} + 1} \tag{6}$$

Taking the modulus of both sides :—

$$\left| \frac{e_o}{e_g} \right| = \frac{-gR_{eq}}{\sqrt{1 + \omega^2 C_s^2 R_{eq}^2}}$$

∴ from equation 3 :—

$$\text{Stage gain at high frequencies} = \frac{\text{Stage gain at medium frequencies}}{\sqrt{1 + \omega^2 \cdot C_s^2 \cdot R_{eq}^2}}$$

As the denominator increases with frequency, this shows that the amplification at high frequencies decreases with increase in frequency.

Equation 6 also indicates that the amplification $\left(\frac{e_o}{e_g}\right)$ is a vector quantity; that is, signals suffer a phase change in passing through the stage. This phase-shift φ is given by :—

$$\varphi = 180° - \tan^{-1} \omega C_s R_{eq}. \tag{7}$$

From this equation, φ appears to be a multi-valued function;

the correct result to take is that for which $\tan^{-1}\omega C_1 R_{eq}$ lies between 0° and 90°; because, as ω decreases, so φ approaches 180°, which is the value of the phase-shift in the stage for medium frequencies.

It is customary to disregard this 180°, and to say that over the medium frequency range there is no phase-shift, and that for high frequencies there is an advance φ' in phase relative to medium frequencies, given by :—

$$\varphi' = -\tan^{-1}\omega C_1 R_{eq} \tag{8}$$

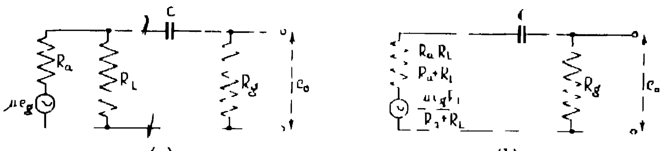


FIG. 355.—Equivalent circuits of an R-C coupled stage at low frequencies

Similarly, Thévenin's theorem can be applied to the low-frequency equivalent circuit of Fig. 355*a*, the result being the circuit *b*.

From this it follows that :—

$$e_o = \frac{\mu e_g R_L}{R_a + R_L} \cdot \frac{R_g}{\dfrac{R_a R_L}{R_a + R_L} + R_g - \dfrac{j}{\omega C}}$$

Hence $$\frac{e_o}{e_g} = -\frac{\mu R_L R_g}{\left(R_a + R_L\right)\left(R - \dfrac{j}{\omega C}\right)}$$

where $$R = \frac{R_a R_L}{R_a + R_L} + R_g$$

This may be re-arranged thus :—

$$\frac{e_o}{e_g} = \frac{-\mu R_L \cdot R_g}{(R_a + R_L)} \cdot \frac{1}{\left(R - \dfrac{j}{\omega C}\right)}$$

$$\frac{e_o}{e_g} = \frac{-\mu \cdot R_L \cdot R_g}{R_a R_L + R_a R_g + R_L R_g} \cdot \frac{1}{\left(1 - \dfrac{j}{\omega C R}\right)}$$

$$= \frac{-g \cdot R_{eq}}{1 - \dfrac{j}{\omega C R}} \tag{9}$$

Taking the modulus of both sides :—

$$\left|\frac{e_o}{e_g}\right| = \frac{-g R_{eq}}{\sqrt{1 + \left(\dfrac{1}{\omega C R}\right)^2}}$$

∴ from equation 3 :—

$$\text{Stage gain at low frequencies} = \frac{\text{Stage gain at medium frequencies}}{\sqrt{1 + \left(\frac{1}{\omega CR}\right)^2}} \qquad (10)$$

Since $\frac{1}{\omega C}$ increases with decrease in frequency, this shows that the stage gain at low frequencies decreases with decrease in frequency.

Also from equation 9 the phase-shift at low frequencies is $\tan^{-1}\left(-\frac{1}{\omega CR}\right)$ relative to the phase-shift at medium frequencies. With the convention of zero phase-shift at medium frequencies, the phase-shift at low frequencies is therefore:—

$$\varphi' = \tan^{-1}\left(\frac{1}{\omega CR}\right) \qquad (11)$$

Thus equation 11 indicates that at low frequencies there is a positive phase-shift (relative to the phase at medium frequencies), and this phase-shift increases as the frequency decreases; and equation 8 indicates that there is a negative phase-shift at high frequencies which increases with increased frequency.

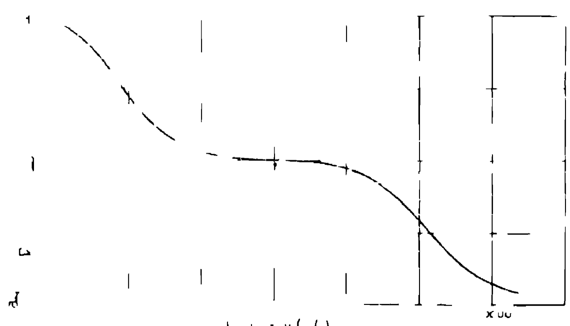


Fig. 356. Phase-frequency characteristic of a typical R-C coupled stage

For a single R-C coupled stage the phase-shift will never exceed 90° either way at any frequency. A graph showing a typical phase-frequency characteristic (the phase-shift being measured relative to middle frequencies) is shown in Fig. 356.

Universal response curves for R-C coupled amplifiers

It is often convenient to have a response curve that may be applied to all R-C coupled amplifiers. Considering first the low frequencies. Let f_0 be the frequency at which the reactance

$X_c\left(=\frac{1}{\omega C}\right)$ of the coupling condenser is numerically equal to the resistance $R = \frac{R_a R_L}{R_a + R_L} + R_g$. This frequency comes within the range of low frequencies.

Then, from equation 10, the stage gain at f_o will be 0·707 times the stage gain at medium frequencies; and from equation 11, the phase shift at f_o will be $+45°$ relative to the phase-shift at medium frequencies. In the same way the stage gain and phase-shift may be found for any other low frequency expressed as a multiple of f_o. Table XIV is prepared in this way.

TABLE XIV
Gain of R-C amplifier at low frequencies

Frequency	Relative amplification (voltage ratio)	Relative amplification in decibels = 20 . $\log_{10}$ (voltage ratio)	Relative phase-shift
$0·1\,f_o$	0·100	-20·0	84° 18′
$0·2\,f_o$	0·196	—14·2	78° 42′
$0·5\,f_o$	0·447	7·0	63° 26′
f_o	0·707	3 0	45°
$2\,f_o$	0·895	0·97	26° 34′
$5\,f_o$	0·980	0·17	11° 18′

Now considering the high frequencies, let f_o' be the frequency at which the reactance of the total shunting capacity C_t is numerically equal to the resistance $R_{eq} = \cfrac{1}{\cfrac{1}{R_a} + \cfrac{1}{R_L} + \cfrac{1}{R_g}}$

TABLE XV
Gain of R-C amplifier at high frequencies

Frequency	Relative amplification (voltage ratio)	Relative amplification in decibels = 20 $\log_{10}$ (voltage ratio)	Relative phase-shift
$0·2\,f_o'$	0·980	— 0·17	-11° 18′
$0·5\,f_o'$	0·895	— 0·97	—26° 34′
f_o'	0·707	— 3·0	—45°
$2\,f_o'$	0·447	— 7·0	—63° 26′
$5\,f_o'$	0·196	-14·2	—78° 42′
$10\,f_o'$	0·100	—20·0	—84° 18′

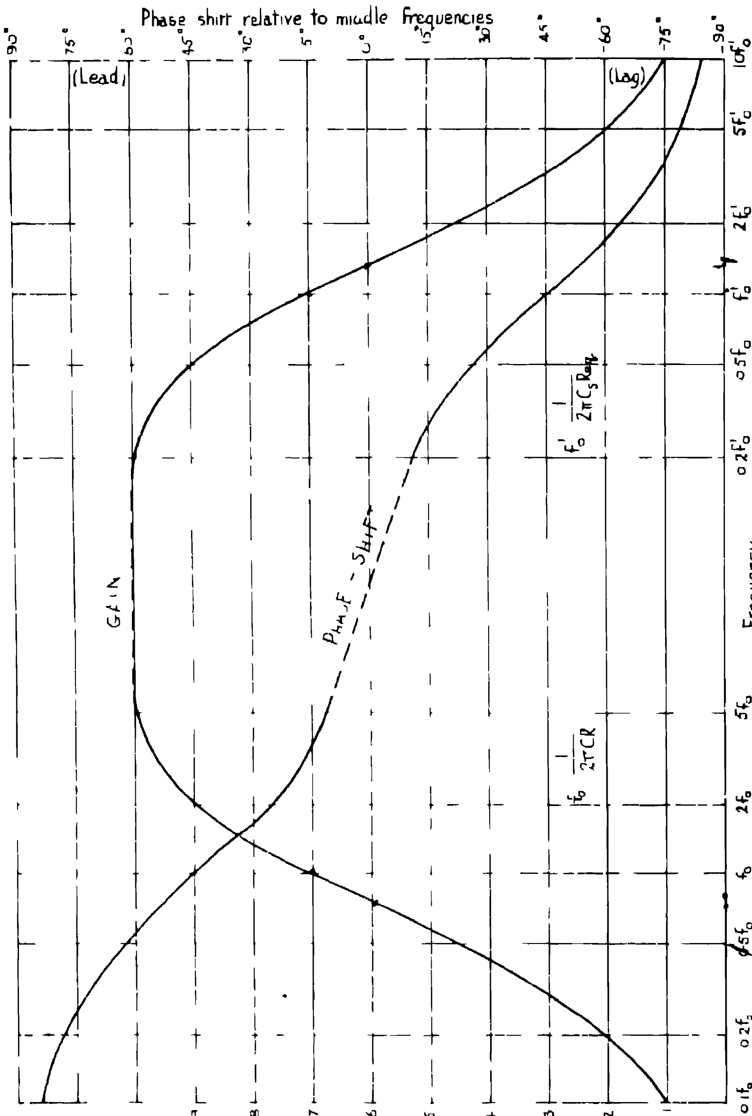


Fig 357 — Universal response curves for a resistance-capacity coupled amplifier.

Then, from equation 7, the stage gain at f_o' is 0·707 times the stage gain at medium frequencies, and from equation 8 the phase shift at f_o' is $-45°$ relative to that at medium frequencies. Table XV is compiled for other high frequencies expressed in terms of f_o'.

Fig 357 shows a universal response curve for a resistance-capacity amplifier. The procedure for obtaining the response curves for a particular amplifier is as follows:

(i) Calculate the stage gain at middle frequencies using equations 1 or 3.

(ii) Calculate the frequency f_c from the relationship

$$f_c = \frac{1}{2\pi CR}$$

and read off the gain and phase shift at low frequencies from the curves of Fig 357.

(iii) Estimate the total shunting capacity C_s and calculate the frequency f_o from the relationship

$$f_o = \frac{1}{2\pi C_s R_{eq}}$$

then read off the gain and phase-shift at high frequencies from the curves of Fig 357.

Example.—

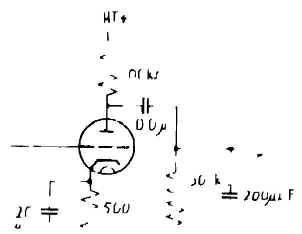


Fig. 358. R-C coupled stage to illustrate use of universal response curve.

Consider a circuit (*see* Fig 358) having the following constants:—

μ	100	R_a	100 000Ω
R_L =	200 000Ω	R_g	500 000Ω
C_s =	200 μμF	C =	0·01μF

Find the mid-frequency gain and those frequencies at the top and bottom of the range at which the gain has dropped by 6 db.

Step 1.—Mid-frequency gain

$$R_L' = \frac{R_L R_a}{R_L + R_a} = 143{,}000\,\Omega$$

$$\text{Gain} = \frac{\mu R_L'}{R_L' + R_a} = 100 \cdot \frac{143{,}000}{243{,}000} = 58{\cdot}8$$

Step 2.—Low-frequency gain

$$f_o = \frac{1}{2\pi CR}, \text{ where } R = \frac{R_L \cdot R_a}{R_L + R_a} + R_g = 567{,}000\,\Omega$$

$$\therefore \quad f_o = \frac{10^8}{2\pi \cdot 567{,}000} = 27 \text{ c/s}$$

But f_o is the frequency at which the gain is 0·707 times its mid-frequency value. The frequency required is that at which the gain has dropped by 6 db; *i.e.*, the voltage gain has dropped to half its mid-frequency value. This can be found from Fig. 357, and is equal to about $0{\cdot}6\,f_o$, *i.e.* about 16 c/s.

Step 3.—High frequency gain

$$f_o' = \frac{1}{2\pi C_s R_{eq}}, \text{ where } R_{eq} \quad (R_a, R_L, \text{ and } R_g \text{ in parallel})$$

$$R_L' = (R_L \text{ and } R_g \text{ in parallel}) = 143{,}000\,\Omega$$

$$\therefore \quad R_{eq} = \frac{143{,}000 \cdot 100{,}000}{243{,}000} = 58{,}800\,\Omega$$

$$\therefore \quad f_o = \frac{10^{12}}{2 \cdot \pi \cdot 200 \cdot 58{,}800} = 13{,}500 \text{ c/s}$$

From Fig. 357, the frequency at which the gain is half the mid-frequency gain is about $1{\cdot}7\,f_o$, *i.e.* about 23,000 c/s.

Load line for R-C coupled amplifiers

It has been seen that the load line for a valve with a resistance R_L in the anode circuit passes through the HT supply voltage point

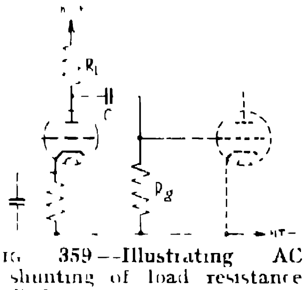


FIG. 359.—Illustrating AC shunting of load resistance R_L by R_g

on the horizontal axis. In the circuit of Fig. 359, this is true only as far as direct currents are concerned; the DC load line will pass through E_b, and will have a slope $\frac{1}{R_L}$, and the operating point of the valve will lie on this line.

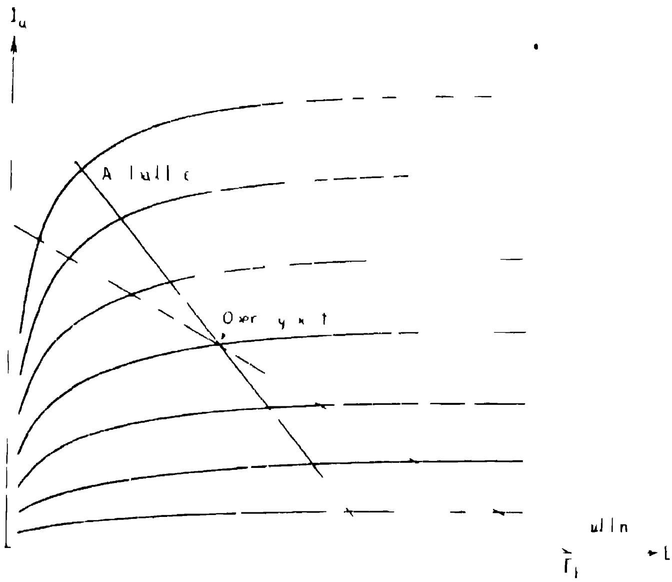


FIG. 360.—DC and AC load lines

For rapid variations of anode voltage, however, C has a low impedance (in the working frequency range) and the equivalent load on the valve R_I is equal to (R_L and R_g in parallel) i.e. $\dfrac{R_L R_g}{R_L + R_g}$. So far as AC variations are concerned, therefore, the load line has a slope of $\dfrac{R_L + R_g}{R_L R_g}$, and passes through the operating point. This will not be identical with the DC load line (whose slope is $\dfrac{1}{R_L}$) unless $R_g \gg R_L$. The two load lines are shown in Fig. 360.

CHOKE AND TRANSFORMER-COUPLED VOLTAGE AMPLIFIERS

Choke-capacity coupling

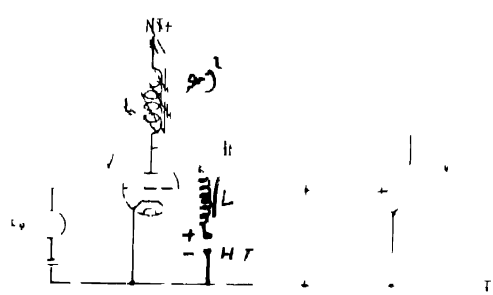


FIG. 361.—A simple choke capacity coupled stage

Fig 361 shows the simple circuit of a choke-capacity coupled amplifier stage A comparison with Fig 349*a* shows that this differs from resistance capacity coupling only in that the anode load resistor R_L is replaced by an iron-cored inductance L This has the advantage that, although it offers a high impedance to the alternating components of the anode current it can be made to have a low DC resistance The DC voltage drop across it will be small, and under static conditions practically the whole available HT voltage will be applied to the anode, thus enabling lower voltage HT supplies to be used The frequency response is reasonably flat over a central range of frequencies, but at low frequencies the gain falls off because of the decreased impedance of the choke at low frequencies In order to extend the flat response to lower frequencies it is necessary to have as high a value of inductance

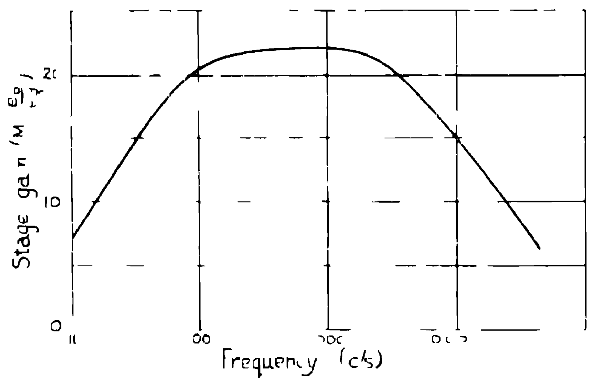


Fig. 362 Gain frequency response of a typical choke capacity coupled stage

as possible so that the impedance of the choke is still quite large at low frequencies thus to give an impedance of 60 kΩ at 500 c/s the choke would require an inductance of about 20 henries The coupling condenser also causes a falling off of gain at low frequencies, just as in the RC coupled amplifier, but its capacity can always be made so large that the falling off in gain at low frequencies due to the decrease in load impedance occurs considerably before that due to excessive condenser reactance

As the frequency is increased the gain tends to rise due to the increase in the reactance of L and the reduction in the reactance of C with increase in frequency The rise in gain continues until frequencies are reached at which the reactance of C may be neglected and the reactance of L is very large compared with R_g and R_a in parallel The value of R_g affects not only this gain, at medium frequencies but also the range over which the gain is substantially flat At high frequencies the gain drops off due to the shunting effect of the input capacity of V_2 Fig 362 shows a typical gain-frequency response curve for a choke-capacity coupled stage.

Load line for choke-capacity coupled amplifier

A reactive load on a valve is represented on the characteristics as an elliptical load line, and not a straight line. This follows from

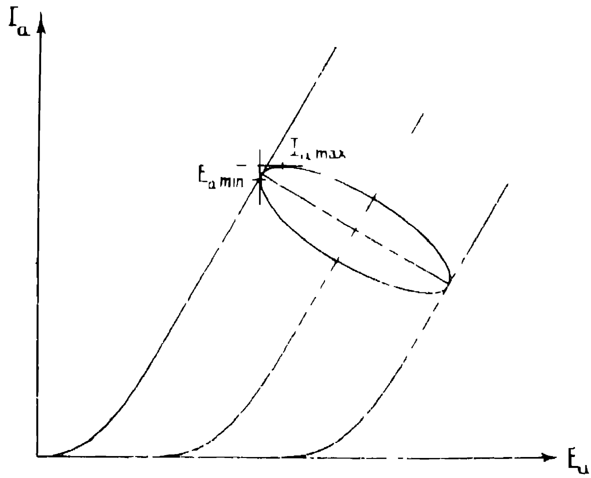


FIG. 363.—Reactive load ellipse

the fact that, with a reactive load, the anode voltage and current are out of phase, and so peaks occur at different instants. The load line, being the locus of points representing the current and voltage at different instants, can be shown from the equations for E_a and I_a to be, in fact, an ellipse, as shown in Fig. 363.

Transformer coupling

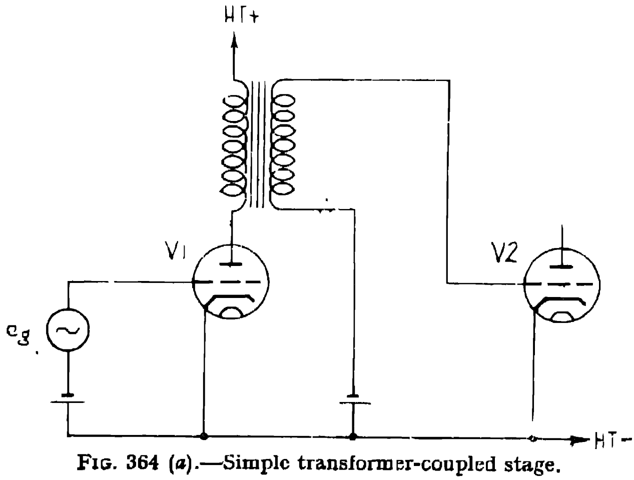


FIG. 364 (*a*).—Simple transformer-coupled stage.

Fig 364*a* shows a transformer-coupled amplifier; it will be seen that the primary of the transformer forms the load impedance and the secondary is connected between grid and cathode of the following stage.

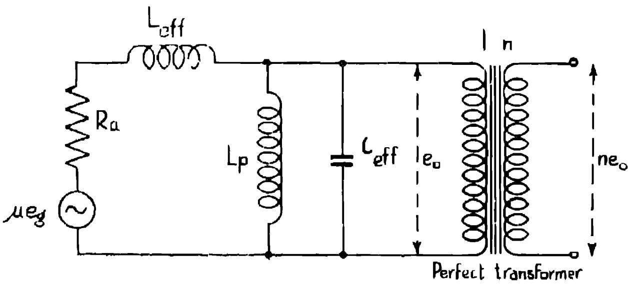


FIG. 364 (*b*) — Simplified equivalent circuit of transformer coupled stage.

Fig 364*b* shows a simplified equivalent circuit; compare this with Fig 207 (of Chapter 5).

L_p = primary inductance of transformer.

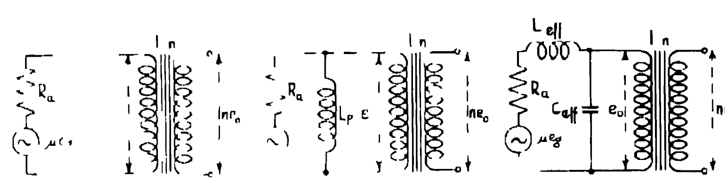
$L_{eff} = L_1 + \frac{1}{n^2} L_2$ = total leakage inductance, referred to primary.

L_1 = leakage inductance of primary winding.

L_2 = leakage inductance of secondary winding.

C_{eff} = shunt capacity made up of the self-capacity of the primary winding and of the self-capacity of the secondary and the input capacity of V_2 both referred to the primary.

The full equivalent circuit would also contain components to account for primary and secondary resistance, eddy current and hysteresis losses and capacity between the windings, but all these are small enough in practice to be neglected and have been omitted for simplicity.



(*a*) (*b*) (*c*)

FIG. 365 Simplified equivalent circuits of transformer-coupled stage applicable to medium low and high frequencies

Fig 365 gives the simplified equivalent circuits relating respectively to the middle low and high ranges of the frequency band. Over the middle range (*see* Fig. 365*a*) the impedance of the primary inductance is high (L_p may be about 20 henries), and the

reactances of L_{eff} and C_{eff} will be small. They can therefore be neglected, and the stage gain will be —

$$M = \left|\frac{ne_o}{e_g}\right| = n\left|\frac{e_o}{e_g}\right| = n\mu \qquad (12)$$

One advantage of transformer coupling therefore is that by using a step-up transformer it is possible to obtain a stage gain exceeding the amplification factor of the valve.

At low frequencies the reactance of the transformer primary will be lower and must be considered whilst the reactances of C_{eff} and L_{eff} are still negligible. The equivalent circuit is then that shown in Fig. 365b. In this case

$$M = \left|\frac{ne_o}{e_g}\right| = n\left|\frac{e_o}{e_g}\right| = n\frac{\omega L_p\mu}{\sqrt{R_a^2+\omega^2L_p^2}}$$

i.e.

$$M = \mu n\frac{1}{\sqrt{1+\left(\frac{R_a}{\omega L_p}\right)^2}} \qquad (13)$$

This shows that the stage gain falls as the frequency is decreased, the rate at which it falls being dependent on $\frac{R_a}{L_p}$. If L_p is made larger the stage gain will remain reasonably constant down to a lower frequency. For a wide frequency range working down to low frequencies a very high value of primary inductance is necessary.

At high frequencies the reactances of L_{eff} and C_{eff} become important and the circuit of Fig. 365c applies. Then

$$M = \left|\frac{ne_o}{e_g}\right| = n\left|\frac{e_o}{e_g}\right| = \frac{n\mu\frac{1}{\omega C_{eff}}}{\sqrt{R_a^2+\left(\omega L_{eff}-\frac{1}{\omega C_{eff}}\right)^2}} \qquad (14)$$

Now suppose the series resonant frequency of L_{eff} and C_{eff} is ω_o,

i.e.

$$\omega_o L_{eff} = \frac{1}{\omega_o C_{eff}} \qquad (15)$$

and let

$$x = \frac{\omega}{\omega_o} \qquad (16)$$

Substituting $\omega = x\omega_o$ in equation 14

$$M = \frac{\mu n\frac{1}{x\omega_o C_{eff}}}{\sqrt{R_a^2+\left(x\omega_o L_{eff}-\frac{1}{x\omega_o C_{eff}}\right)^2}}$$

i.e.

$$M = \frac{\mu n}{\sqrt{x^2R_a^2\omega_o^2C_{eff}^2+(x^2\omega_o^2L_{eff}C_{eff}-1)^2}}$$

i.e.
$$M = \frac{\mu n}{\sqrt{\left(\frac{x}{Q_0}\right)^2 + (x^2 - 1)^2}} \tag{17}$$

where
$$Q_0 = \frac{1}{R_a \omega_0 C_{eff}} = \frac{\omega_0 L_{eff}}{R_a} \tag{18}$$

i.e. Q_0 is the circuit Q at the frequency for which C_{eff} and L_{eff} are in series resonance.

By differentiating equation 17 it can be shown that the stage gain is maximum with respect to x when —

$$x^2 = 1 - \frac{1}{2Q_0^2} \tag{19}$$

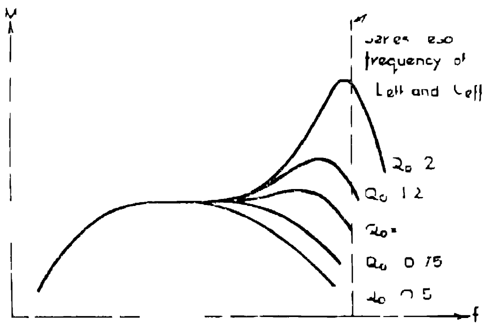


Fig. 366. Gain-frequency response of transformer coupled stage showing effect of Q_0 on high frequency response.

Fig. 366 shows typical response curves for a transformer coupled stage, and the effect on gain at high frequencies of a variation in Q_0.

From the curves it is clear that the flattest response is obtained when Q_0 is slightly less than unity. This means that for a given transformer there is a proper value of R_a which must be used to give the correct value of Q. Intervalve transformers are usually produced to work with valves of AC resistance of about 10 000 ohms, i.e. with a triode. Intervalve transformer coupling is not often used with pentodes over the audio range, but where it is used the transformer primary must be shunted by a resistance (see Fig. 367).

The simplified equivalent circuit is shown in Fig. 368a; this clearly reduces (by Thevenin's theorem) to that of Fig. 368b.

R_{eq} is the equivalent resistance of R_a and R in parallel, and since, for a pentode, R_a is very large, $R_{eq} \simeq R$. At high frequencies therefore R determines the value of Q_0 and hence the high frequency response; the value of R selected is thus the value giving the optimum circuit Q at the resonant frequency of L_{eff} and C_{eff}.

7. One requisite of an intervalve transformer for use in the circuits

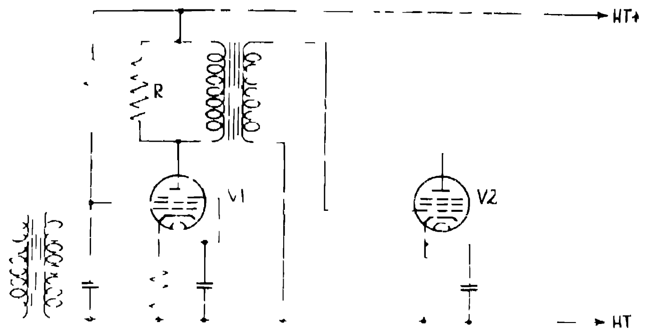
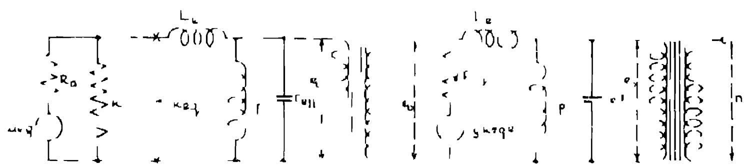


FIG. 367.—Transformer coupled stage using pentode.

far considered is that it shall have a large value of primary inductance and at the same time shall be capable of carrying the standing anode current in its primary without undue saturation of the iron core. If saturation occurs, distortion will be produced.



(a) (b)

FIG. 368.—Equivalent circuit of transformer coupled stage using pentode.

In addition the value of the inductance of the transformer windings will drop. This places practical limitations on the transformer, making it bulky and of low step-up ratio. By using a choke parallel feed circuit as in Fig. 369 these difficulties are largely

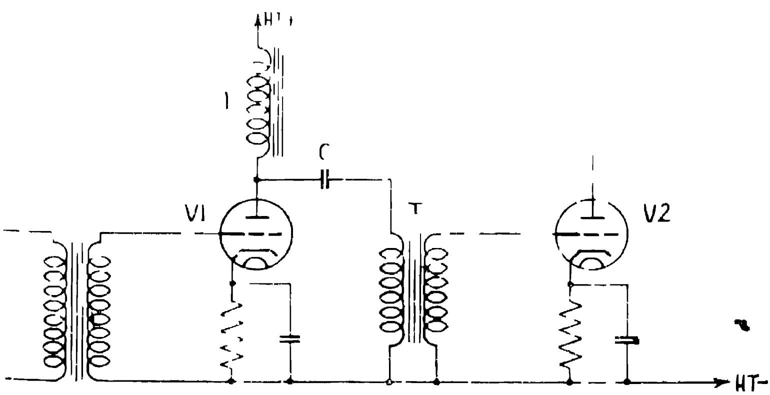


FIG. 369.—Transformer coupling using parallel feed.

overcome; the resulting transformer can be made more compact, with a core of very high permeability material. Comparing two transformers of the same size, one designed for parallel feed and the other for series feed: the parallel-feed transformer can be made to have about three times the primary inductance, and roughly the same value of leakage inductance, as the series-feed transformer. The choke can generally be made to have a larger inductance than a transformer for use in the same circuit, for it has only one winding and iron losses are less important. The overall result is a flatter gain-frequency response at low frequencies, although there will still be a falling-off at very low frequencies produced by the choke and coupling condenser.

Input transformers

Where it is necessary to couple the grid of the first stage of an amplifier to a low impedance such as a transmission line, this is usually done by means of a step-up transformer as in Fig. 370.

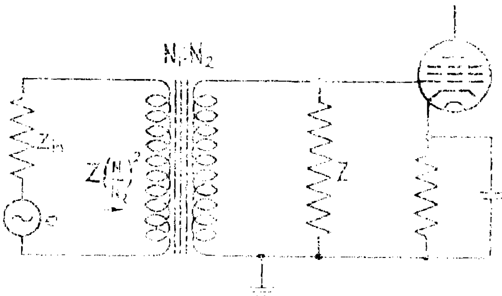


FIG. 370.—Input transformer.

It will be seen that this is, in all essential features, equivalent to the interstage coupling of Fig. 364 with the input impedance of the line replacing the AC resistance of V_1. Thus the equivalent circuit of Fig. 364, and the analysis which follows, holds for input transformers as well as interstage transformers. The primary inductance will be in proportion to the impedance of the line and the transformer may have a high step-up ratio (*e.g.* 1 to 22). In order to obtain the correct input impedance, an impedance Z may be placed across the secondary of the transformer.

TUNED VOLTAGE AMPLIFIERS

In the commonest type of tuned amplifier the load impedance is supplied by a parallel resonant circuit, which gives the necessary high impedance load over a comparatively narrow band of frequencies. Such amplifiers can be made very selective with respect to frequency, making it possible to amplify signals of a desired frequency whilst eliminating other signals.

Amplifiers of this type are used

(a) for amplification of radio frequency signals;

(b) for amplification of intermediate frequencies in super-heterodyne radio frequency receivers;

(c) for amplification of [illegible] frequencies within the audio range [illegible] in VF [illegible] equipment;

(d) for amplification of 500 c/s voice frequency (VF) signals in many types of VF signalling unit.

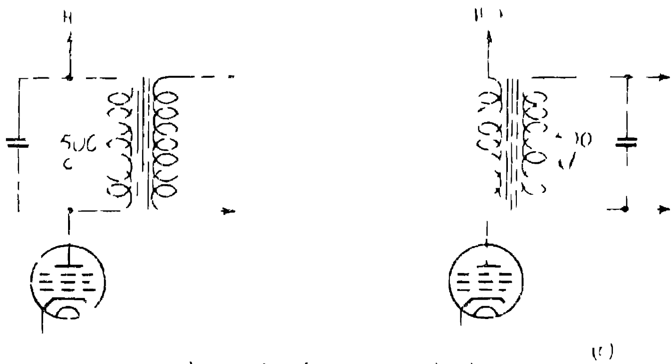


Fig. 371.—[illegible]

Fig. 371 shows two [illegible] circuits of the tuned anode load which is the basis of many VF signalling units.

In both cases the response curve is of the type shown in Fig. 372.

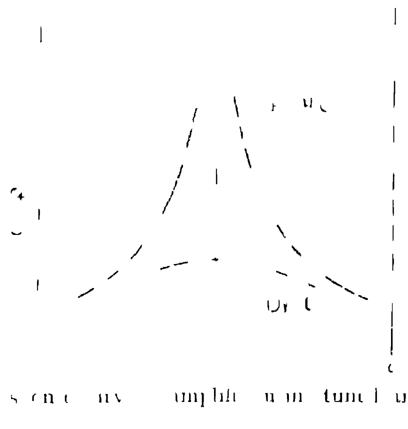


Fig. 372.—Response curves of amplifiers with tuned anode loads.

This shows that if the Q of the parallel resonant circuit is high, the amplification at resonance will be high and the frequency discrimination will be critical. This is the ideal response curve

for a VF signalling receiver where it is desired to select only one frequency, but if it is required to have a constant gain over a band of frequencies, say from 1800 to 1920 c/s, it is necessary to reduce the Q of the tuned circuit. This will decrease the amplification at resonance and may necessitate an additional stage of voltage

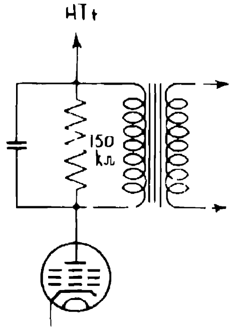
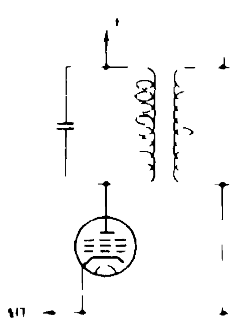


Fig. 373. Tuned amplifier with shunt resistance to broaden the frequency response

amplification in order to obtain the required gain. The simplest method of reducing the Q is the introduction of a shunt resistance across the tuned circuit as shown in Fig. 373.

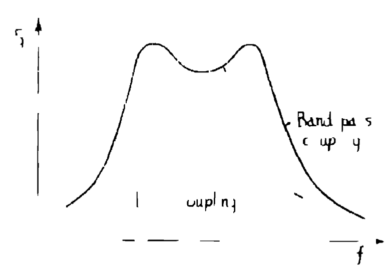
Band-pass coupling

For intermediate frequency (IF) amplification in a superheterodyne receiver a different form of coupling is used between stages. This is *band-pass coupling* as shown in Fig. 374.



(a)

Fig. 374. Tuned amplifier with band-pass coupling



(b)

Fig. 375. Response curve of tuned amplifier using band-pass coupling

Here both primary and secondary of the transformer are tuned to the required intermediate frequency (say, 465 kc/s) and, with suitable values of Q and of the coupling coefficient, the response curve will be as in Fig. 375.

Clearly such an amplifier gives very good discrimination against frequencies appreciably off resonance while at the same time it responds uniformly to a band of frequencies immediately above and below the resonant frequency. Two such IF transformers may be coupled together as in Fig. 376, the response curve being

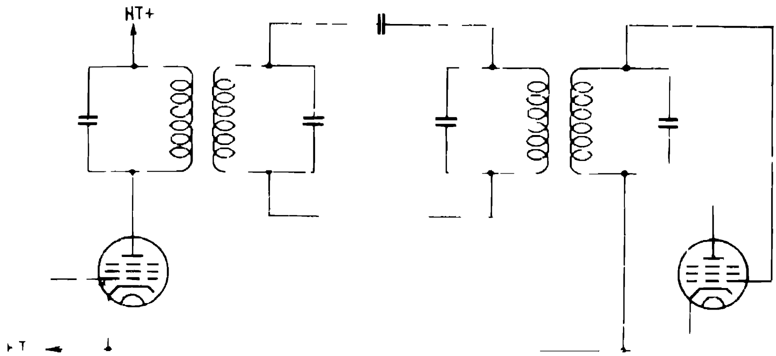


Fig. 376.—Typical coupling of two transformers using [illegible] coupling

similar to Fig. 375 but giving greater discrimination against frequencies appreciably off resonance.

Shunt resistances may be inserted across primary or secondary to flatten the response if required. At the same time this will reduce the frequency discrimination.

GAIN CONTROL IN AMPLIFIERS

A large number of amplifiers used in line terminal equipment utilise a method of gain control depending on the application of varying amounts of negative feed back. The principles of this method are dealt with in the next chapter. Other methods that

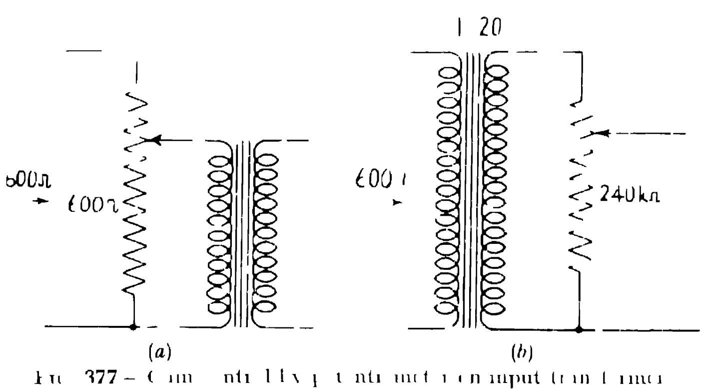


Fig. 377.—Gain control by potentiometer on input transformer

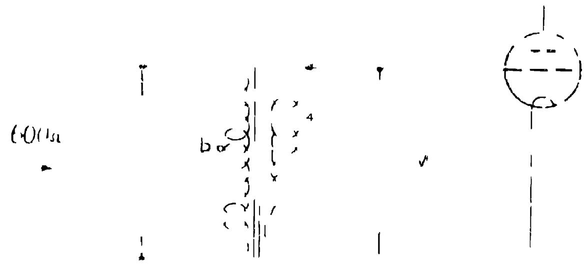
will be encountered either singly or in combination, for producing variation in gain, are

(i) A constant gain amplifier is provided with variable attenuation pads in the input circuit, the output circuit or both.

(ii) A potentiometer is provided on either the low impedance (line) side or the high impedance side of the input transformer.

Fig. 377 shows the [illegible] potentiometer on the low and high impedance sides of the input transformer.

(iii) Variation [illegible] is provided by means of varying tappings on the input transformer thereby varying the turns ratio.



[illegible]

Fig. 378 shows [illegible] both the [illegible] primary [illegible] the secondary turns [illegible] For [illegible] the low impedance [illegible] and the [illegible] on the high impedance [illegible] of ½ db each.

Automatic gain control (AGC)

Automatic gain control [illegible] automatic volume control or AVC [illegible] in radio receivers. The requirement [illegible] a device that will maintain the output [illegible] desired level in spite of [illegible] the [illegible] level of the incoming [illegible] in the case of line communication [illegible] input level [illegible] and generally take place [illegible] than [illegible] in a radio receiver. In [illegible] overall input attenuation can be tolerated but [illegible] system for example [illegible] the operation of the receiver [illegible] the incoming signal unless AGC [illegible] amplifier.

The general principle of AGC is that a portion of the incoming signal after amplification [illegible] rectifier and the [illegible] across a resistor the DC component of the voltage which will be proportional to the amplitude of the incoming signal is utilised to provide negative

bias for the amplifying stages. These stages usually employ variable mu valves. Thus the greater the input signal, the greater [illegible] overall [illegible] applied [illegible] AGC [illegible] however [illegible]

[illegible] in which AGC [illegible]

Examples of AGC circuits

FIG. 379 [illegible] AGC [illegible] the output

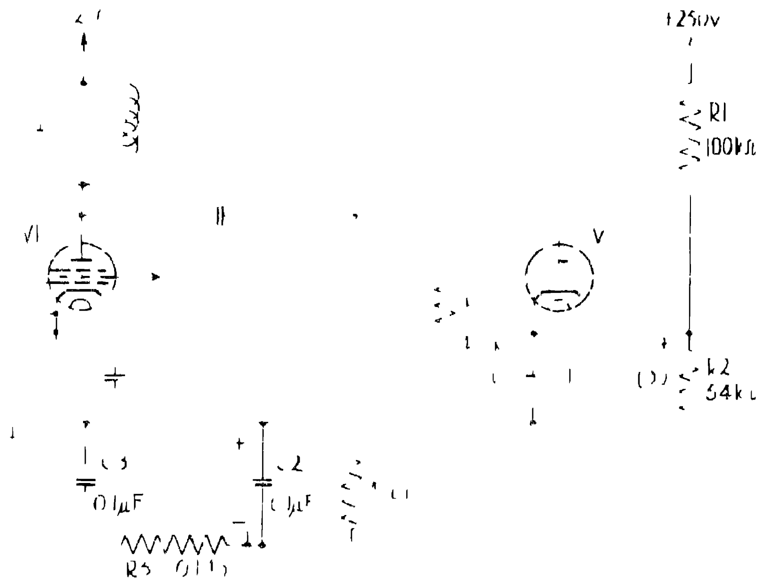


FIG. 379 [illegible]

of the 465 kc/s IF amplifier [illegible] the AGC amplifier valve V_1, which [illegible] tuned amplifier [illegible] similar to those already discussed. The output [illegible] coupled to the diode V_2. The cathode of V_2 is maintained at about 90 volts positive with respect to the earth line by means of the potentiometer R_1 R_2 between HT positive and earth. This provides the voltage delay in the AGC, for no current can flow through the diode until the peak value of the signals applied between anode and earth exceeds 90 volts. When the signals are above this level V_2 acts as a shunt diode

rectifier and produces a rectified component giving a potential across R_3 of polarity as shown in Fig 379 The negative potential of the top end of R_3 relative to earth is applied, via the resistor R_4 (decoupled by C_2) and *via* individual smoothing circuits such as R_5 and C_3 to the grids of the variable mu pentodes of the preceding amplifier stages Since these amplifiers have tuned anode loads, the output is practically free from any harmonic distortion that may be produced in the valve due to working on the curved portion of the mutual characteristic

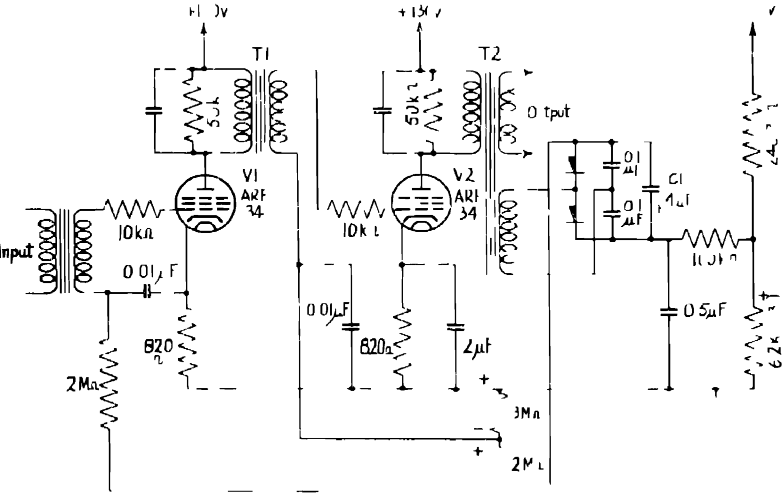


Fig 380 AGC circuit used in VF telegraph equipment

Fig 380 shows another AGC circuit this time used in the receive side of a VF telegraph equipment Part of the output of a tuned amplifier is applied to a voltage doubler circuit that develops a DC potential across C_1 as shown The AGC bias is applied to V_1 and V_2 via a potential divider that gives V_1 a larger AGC bias than V_2 Delay voltage is provided by a potentiometer (6 2 kΩ and 246 2 kΩ) between HT positive and earth If the incoming signal increases in amplitude the negative bias voltage produced by the AGC circuit and applied to V_1 and V_2 will be increased and the gain of the amplifier reduced If the incoming signal decreases in amplitude the negative bias will decrease and the gain of the amplifier increase In this way the output voltage across the secondary of the transformer T_2 is kept substantially constant even though large changes may occur in the input level In this case the AGC is so effective that a decrease in the input voltage by as much as 900 times (59 db) may have no appreciable effect on the output voltage This enables the apparatus concerned to operate satisfactorily on all input voltages from 1 8 volts down to 2 millivolts without any manual gain control

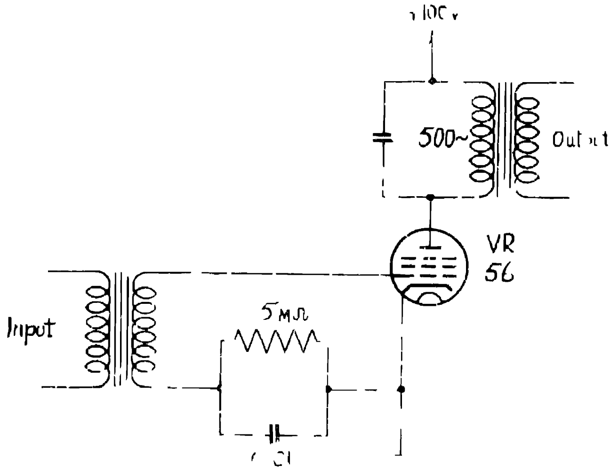


Fig. 381.—Limiter amplifier used in a V.F. signalling receiver

Another circuit that may be classed as giving a type of AGC is shown in Fig. 381. This is the first stage of a V.F. signalling receiver and its function is to give a constant voltage output independent of the amplitude of the 500 c/s input provided that the input exceeds a certain minimum level. It will be seen that it consists merely of a tuned amplifier to which grid leak bias is

Fig. 382.—Action of limiter amplifier

applied. It has already been seen that when grid leak bias is applied to a valve, the bias voltage on the grid is just sufficient to cause a small amount of grid current to flow on the positive peaks of applied signal (*see* p. 370).

In Fig. 382 there are shown three signals of varying amplitudes For small signals such as A, the operating point is at a, and the output signal a applied to the tuned circuit in the anode will give a voltage output that varies with the amplitude of the input signal If the amplitude of the input is increased to that of the signal B the operating point recedes to b and the output will be β. That the output now becomes independent of any further increase in the input level is seen from a consideration of the input signal C, which gives the output γ. It is thus the value of the cut-off bias that determines the value of input signal above which limiting takes place, and to make this value small it is usual to supply the limiter valve with a low anode voltage It will be seen that the distortion produced by such a limiter with large input signals is very great.

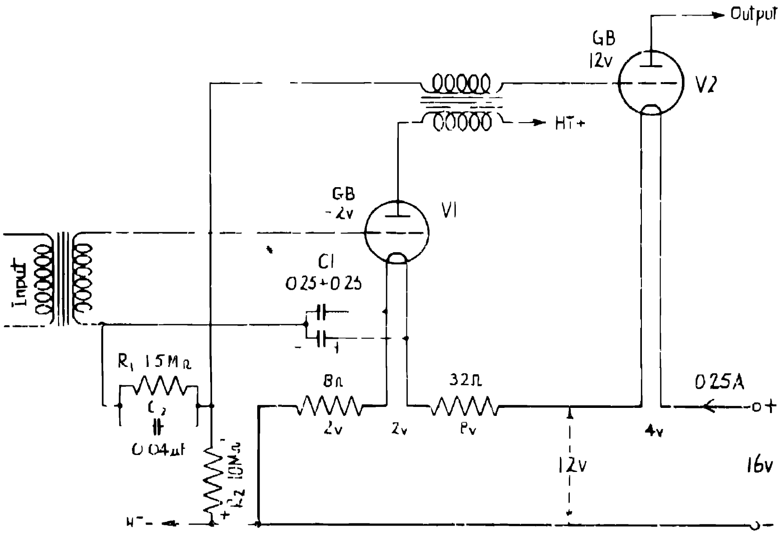


FIG 383 AGC circuit used in a VF telegraph detector.

As a last example of AGC, consider the amplifier-detector of a VF telegraph system (*see* Fig. 383). The standing bias arrangements of this circuit (filament bias) have already been considered. The first stage V_1 has a 2-volt bias, and the second V_2 a 12-volt bias. The first stage has a high gain, and for large signals V_2 tends to run into grid current on the positive peaks. When grid current flows, a potential is developed across R_2, and the two parts of condenser C_1 are charged with polarity as shown, and the

voltage across these condensers is applied to V_1 as AGC bias. The discharge path for C_1 is *via* R_1 and R_2, giving a time constant of 6·5 seconds. In order to prevent the amplifier being cut-off by its AGC bias due to a sudden high level transient, C_2 and R_1, with a time constant of 60 milliseconds, impose a time delay on the application of the AGC bias to V_1; thus the bias on V_1 will not increase until an increase in the level of the received signal has been maintained for a period of that order AGC bias is also applied to V_2, but it has practically no effect on this valve, since the AGC voltage is small compared with the large initial bias.

DISTORTION AND NOISE IN AMPLIFIERS

Distortion

Any transmission system (an amplifier being a particular example) is said to introduce distortion if the input and output signals are not *identical in waveform.* This change in waveform may occur due to one or more of a number of causes. These different types of distortion are rigidly defined below for the sake of reference, although in practice it may be difficult to separate the distortion produced by a particular system into these subdivisions.

Attenuation distortion

The name "attenuation distortion" is applied to the case of a transmission system where there is a variation of gain or loss with frequency It is assessed with the system operated under steady-state conditions by applying a series of signals of sinusoidal waveform at different frequencies.

Phase distortion

Phase distortion occurs when the time of propagation through a transmission system varies with frequency. Owing to the different relative phase relationships then existing, the output waveform may appear to be quite different from the input waveform, even though the same frequencies are present in the same relative amplitudes.

Phase distortion will always be present unless the graph of the overall phase-shift plotted against frequency is a straight line passing through a point on the phase-shift axis corresponding to zero or some integral multiple of 2π radians.

It may be noted that the phase distortion encountered in an audio frequency amplifier is, in general, not important, since the ear is insensitive to small differences in phase.

Non-linear distortion

"Non-linear distortion" is the general name given to a certain type of distortion that occurs when the transmission properties of a system are dependent on the instantaneous magnitude of the applied signal, and it may be sub-divided into amplitude distortion, harmonic distortion, and intermodulation distortion.

Amplitude distortion is defined as the variation of gain or loss of a system with the amplitude of the input. It is measured with the system operated under steady-state conditions with an input of sinusoidal waveform.

Harmonic distortion is due to the production of harmonics in the output when a sinusoidal input of specified amplitude is applied. It is expressed as the ratio of the RMS voltage of all the harmonics in the output, to the total RMS voltage at the output.

Intermodulation distortion is due to the production of combination frequencies in the output when two or more sinusoidal voltages of specified amplitude are applied at the input. For two "parent" frequencies p and q the output may contain frequencies such as $(p \pm q)$, $(2p \pm q)$, $(p \pm 2q)$, *etc.*, in addition to the frequencies p and q.

Noise in amplifiers

All amplifiers give some output even when there is no input signal. Such output is commonly referred to as "noise", a general term that is further sub-divided according to the causes producing it.

Hum

The term "hum" is applied to extraneous output voltages having their origin in associated or adjacent power circuits. The chief causes of hum are the use of AC heater supply, poor smoothing in an HT supply produced by rectification from an AC supply, and pick up due to stray electrostatic and electromagnetic fields produced by adjacent power leads. All modern indirectly heated valves are designed to reduce hum from heater supply to a workable minimum, and a further reduction may be obtained by connecting the centre tap of the heater supply to HT. Hum due to AC ripple in the anode supply can be reduced by providing a more adequate smoothing filter, and that due to stray fields by screening of grid leads, earthing of metal chassis and the use of twisted heater leads. It should be noted that vibrator power packs deriving their power from a DC source may cause strong alternating fields in the neighbourhood of the battery leads which should therefore be twisted.

It is particularly important to suppress as far as possible any hum picked up in the early stages of a multi stage amplifier, since any extraneous voltages produced here will be subject to amplification in the later stages. This statement will apply equally to other types of noise.

Microphonic noise

The term "microphonic noise" is used to cover effects arising from mechanical vibration of parts of the circuit, particularly valves. It may be minimised by mounting the valves in such a way

that they are protected from mechanical vibrations transmitted through the chassis, *e.g.* by using sprung valve holders.

Thermal agitation noise

The random motion of the electrons in a conductor produces minute voltages across the terminals of the conductor, and these voltages are constantly changing in value. This is liable to produce noise, particularly in the earlier stages of an amplifier, where it is subject to further amplification. Since the random motion of the electrons is dependent on the temperature of the conductor, this effect is known as "thermal agitation", and it produces noise voltages that are distributed over the whole frequency spectrum.

Valve noise

Thermionic valves introduce a certain amount of noise into an amplifier, since the anode current is subject to random variations.

Contact noise

Noise may also be caused by poor or intermittent electrical contact. This may arise from dirty or damaged switch contacts, terminals, or connections, leaky condensers, faulty resistances, *etc.* Carbon resistors also are liable to introduce noise, due to changes in contact resistance between adjacent granules. This effect increases with increased current, and precludes the use of carbon resistors as anode loads in the early stages of a high gain amplifier if a low noise level is essential.

POWER AMPLIFIERS

So far, only voltage amplification has been considered; that is to say, the sole aim has been the production of a large *voltage* across the anode load. The case will now be considered where the important factor is the *power* developed in the anode load. Power amplifiers are most conveniently classified according to the conditions under which they operate, as determined by the potentials applied to the grid and anode and by the amplitude of the applied signal voltage on the grid. In general, the operation is the same in the case of wide- and narrow-band amplifiers.

Class "A" power amplifiers

An amplifier is said to operate under "Class A" conditions when the waveform of its output is the same as that of its input, as is the case with the majority of voltage amplifiers. This is achieved by biasing the valve to the centre of the straight portion of the mutual (i_a/e_g) characteristic, and applying an input signal that is small enough not to entail operation off this straight portion.

Triode Valves.—Consider the triode amplifier shown in Fig. **384*a***, and its equivalent circuit given in Fig. **384*b***. As is usual with power amplifiers, the AC anode load for the valve is connected through a transformer; this has negligible DC resistance so the

steady DC anode voltage is equal to E_b for all operating conditions. The transformer turns ratio has been taken as 1 : 1 for convenience; any value might be found in practice.

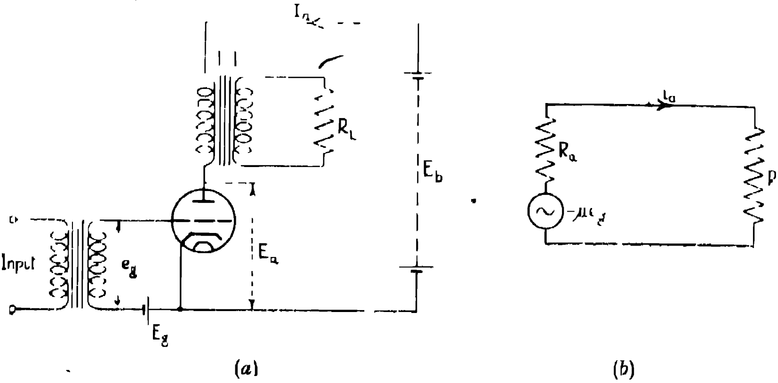


FIG 384 – Simple triode amplifier with equivalent circuit.

The important problem in this circuit is to determine the conditions that will give maximum undistorted *power* in the load R_L for a given supply voltage E_b. It might at first appear, from the maximum power transfer theorem, that the value of R_L giving maximum power output is $R_L = R_a$. This is true only for a given small value of e_g, and under these conditions the power output is low, although it is a maximum with respect to R_L. The absolute maximum power output is obtained by making e_g the maximum permitted by the straight portion of the dynamic characteristic for the particular value of R_L. Under these conditions, the only quantities that can be varied are thus the anode load R_L and the grid bias E_g. This can be simplified still further; for, given any value of R_L, the optimum value of the grid bias E_g is the one that puts the operating point at the centre of the straight part of the valve characteristic. Hence variations of R_L alone will be considered, it being assumed that the bias E_g and the input e_g are adjusted in each case to the optimum values. The optimum value of R_L is most easily obtained from the anode characteristics of the valve.

Consider Fig 385, which shows the anode characteristics of a triode. It is assumed that all the anode characteristic curves are linear above a fixed minimum anode current I_{min}. Let the anode voltage corresponding to I_{min} for $E_g = 0$ be E_0; so E_0 is also fixed. The slope of the curves is given by $\tan \theta = \frac{1}{R_a}$. The load line for R_L is drawn with the "usable" portion as a solid line; it is bounded by the curved region of the characteristics at the bottom, and by grid current at the top of the range. The operating point Q lies on the line E_b, and is adjusted so that it lies at the centre of the working part of the load line. This is done by adjusting the fixed grid bias.

Before the output power can be calculated, an expression must be found for the peak AC anode current i_a obtained with maximum permissible grid swing. Now e_a is the corresponding peak AC anode voltage and it follows that —

$$e_a = i_a \cdot R_L \tag{20}$$

An equation for i_a may be obtained from the fact that the distance $AC = AB + BC$,

i.e.
$$\begin{aligned} E_b - E_o &= AB + e_a \\ &= BD \cot\theta + e_a \\ &= 2\, i_a \cot\theta + e_a \\ &= 2\, i_a R_a + i_a R_L \end{aligned}$$

$$i_a = \frac{E_b - E_o}{2R_a + R_L} \tag{21}$$

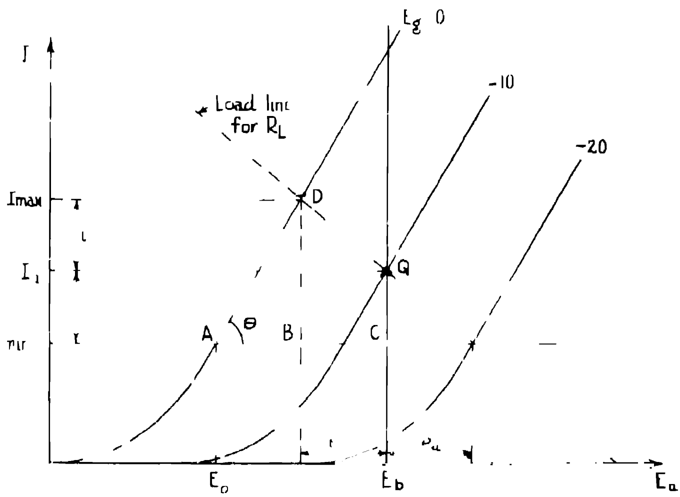


Fig. 355. Anode characteristics of a triode showing load line

The output power can now be calculated —

$P_o = \frac{1}{2} i_a^2 R_L$ (the "$\frac{1}{2}$" is due to i_a being a peak, not RMS, value)

i.e.
$$P_o = \frac{(E_b - E_o)^2}{2} \cdot \frac{R_L}{(2R_a + R_L)^2} \tag{22}$$

The maximum value of P_o will occur when $\dfrac{dP_o}{dR_L} = 0$

i.e., when $\dfrac{(E_b - E_o)^2}{2}\left[\dfrac{1}{(2R_a + R_L)^2} - \dfrac{2R_L}{(2R_a + R_L)^3}\right] = 0$

i.e., when $2R_a + R_L = 2R_L$

i.e., when $R_L = 2R_a$ (23)

Hence the optimum load for a triode valve is twice the AC resistance The maximum output power is —

$$P_{max} = \frac{(I_b \quad E)^2}{2} \cdot \frac{2R_a}{(2R_a + 2R_a)^2}$$

i e

$$P_{max} = \frac{(I_b \quad I)^2}{16 R_a} \qquad (24)$$

It can also be shown that the grid bias is given by —

$$E = \frac{3(I_b \quad I)}{4 \quad \mu} \qquad (25)$$

Pentode and Tetrode Valves With these valves the characteristics are not the same shape as those of a triode and the above results do not apply In this case it is best to draw the anode characteristics and fix an operating point I_b is given

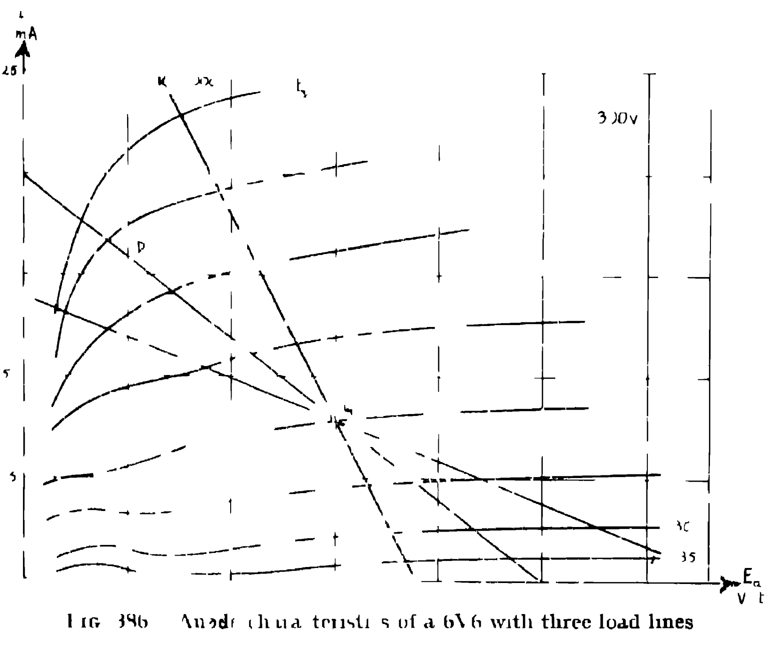


FIG 386 Anode characteristics of a 6V6 with three load lines

and I_a is usually the highest steady current permissible determined by the maximum anode dissipation Various load lines should then be drawn and that one selected which gives *minimum distortion* This will usually be the line that is drawn from the operating point towards the knee of the characteristic

Fig 386 shows the characteristics of a 6V6 valve with three load lines drawn Methods of estimating percentage distortion will be discussed later, but it can be seen that on the 2000-ohm line,

the grid voltage curves become widely spaced at the top and close together at the bottom of the line In the case of the 10,000-ohm line, on the other hand, the curves are cramped at the top. It is clear that the 5000-ohm load affords the best compromise. the grid voltage swing for this load should not exceed the — 5-volt curve.

Efficiency of a Class A power amplifier

In a triode the product of the anode voltage V_a and the standing anode current I_a represent the power dissipated in the valve itself, in the form of heat This is known as the *anode dissipation*, and must not be confused with the output power developed in the load. In tetrodes and pentodes the product of screen voltage and screen current represents an additional power dissipation, *i e.* the *screen dissipation*, the sum of anode and screen dissipation in a pentode or tetrode is called the *total dissipation* The amount of anode (or total) dissipation that can be tolerated is limited by the cooling arrangements within the valve Valve data sheets give maximum permissible values for anode dissipation in the case of output triodes, and for total dissipation in the case of output tetrodes and pentodes

The *efficiency* of a class A power amplifier is given by the ratio of the power output to the power supplied by the HT supply.

Thus :—

$$\text{Efficiency} = \frac{\text{Power output}}{\text{Power supplied by HT supply}}$$

$$= \frac{\text{Power output}}{\text{Power output} + \text{total dissipation}} \qquad (26)$$

Consider a class A power amplifier, and let E_{max} and E_{min}, I_{max} and I_{min}, be the maximum and minimum values of anode voltage and anode current respectively then the RMS value of the alternating voltage is $\frac{E_{max} - E_{min}}{2\sqrt{2}}$, and that of the current is $\frac{I_{max} - I_{min}}{2\sqrt{2}}$.

$$\therefore \quad \text{Power output} = \frac{(E_{max} - E_{min})\,(I_{max} - I_{min})}{8} \qquad (27)$$

For a triode, the power drawn from the HT supply is E_bI_a, where E_b is the voltage of the supply, and I_a is the standing anode current at the operating point

$$\therefore \quad \text{Efficiency} = \frac{(E_{max} - E_{min})\,(I_{max} - I_{min})}{8\,E_bI_a} \qquad (28)$$

Thus it will be seen that this efficiency can never exceed 50 per cent ; for E_{min} and I_{min} can never be less than zero, I_{max} can never be more than twice I_a, and E_{max} must always be less than twice E_b. In practice the efficiency of a triode is never greater than 25 per cent., although in the case of a pentode or beam tetrode the

efficiency may go up to about 35 per cent From equation 26 with this limitation on efficiency clearly the maximum undistorted power output is roughly proportional to the maximum permissible total (anode) dissipation, that is to say if a large power output is required, a valve having a large maximum permissible total dissipation must be chosen

Calculation of harmonic distortion

(*a*) *Triodes* Fig 387 shows a dynamic mutual characteristic for a triode If the operating point Q is chosen in the middle of the straight portion this represents class A working If the working portion of the characteristic is not absolutely straight the output

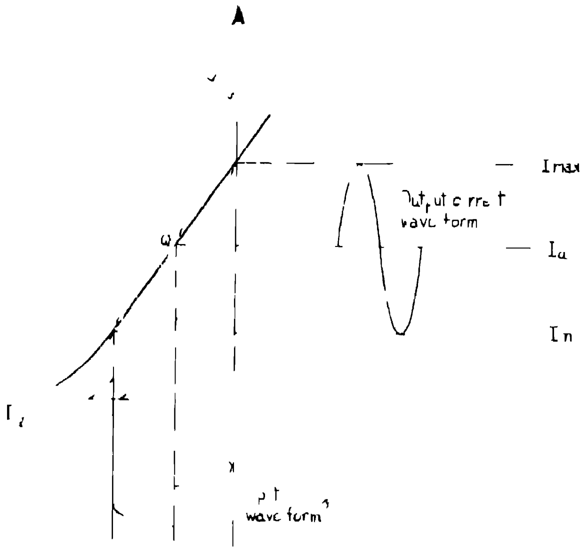


Fig. 387 Dynamic mutual characteristic of a triode showing notation for calculation of harmonic distortion

waveform will not reproduce accurately a sinusoidal input voltage waveform and the result will be harmonic distortion in particular, the output will contain a DC and second harmonic components as well as the fundamental In Chapter 2 it was shown how a Fourier's analysis could be made of an input output curve with the notation of Fig 387, it can be shown that —

$$\text{Direct current component} — A_0 \simeq \frac{I_{max} + I_{min} - 2I_a}{4} \quad (29)$$

$$\text{Fundamental} \quad A_1 \simeq \frac{I_{max} - I_{min}}{2} \quad (30)$$

$$\text{Second harmonic} \quad A_2 \simeq \frac{I_{max} + I_{min} - 2I_a}{4} \quad (31)$$

$$\text{and} \therefore \quad \frac{\text{Second harmonic}}{\text{Fundamental}} = \frac{A_2}{A_1} \simeq \frac{I_{max} + I_{min} - 2I_a}{2(I_{max} - I_{min})} \quad (32)$$

In a triode amplifier, only the second harmonic distortion is important and it is this ratio $\frac{A_2}{A_1}$ which determines the percentage distortion; *i.e.* for maximum undistorted output $\frac{A_2}{A_1}$ must be less than 0·05 to give the arbitrary limit of 5 per cent. distortion which is selected as being the smallest amount of harmonic distortion that can be detected by the human ear.

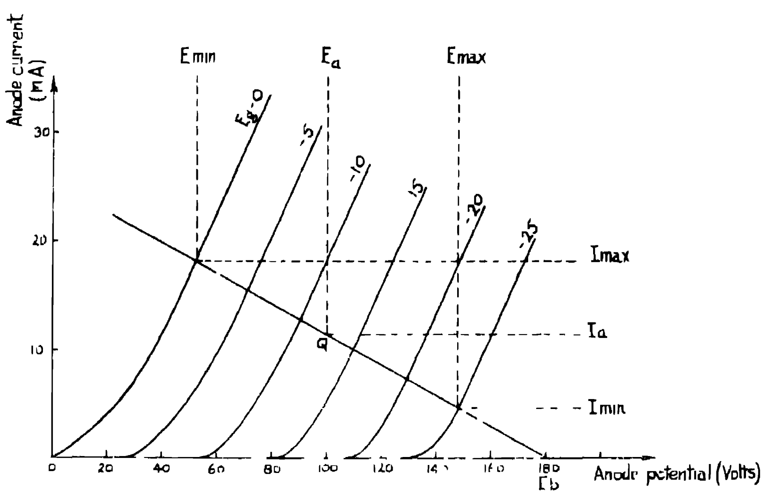


FIG. 388 Load line for a triode, showing notation for calculating harmonic distortion

Fig. 388 shows the same notation applied to the load line on the anode characteristics; equations 29 to 32 also apply in this case.

(*b*) *Pentodes and tetrodes*.—Class A power amplifiers using pentodes or tetrodes require special consideration because the anode current depends on the screen voltage and is substantially independent of the anode voltage. Consequently the mutual characteristics for a constant screen voltage are as shown in Fig. 389, *i.e.* they are practically coincident except at low anode voltages. The dynamic characteristic will therefore be as shown. As compared with the dynamic characteristic of a triode this shows much greater curvature and instead of becoming straighter as the load resistance is increased, develops a point of inflection and the curvature increases. Such a dynamic characteristic gives rise to higher harmonics than second, the main distortion being third harmonic.

From the dynamic characteristic (*see* Fig. 389) the anode current corresponding to certain values of applied signal can be obtained.

Let I_a = anode current for zero signal voltage.

I_{max} = anode current for maximum positive signal peaks

I_{min} = anode current for maximum negative signal peaks.

I_2 = anode current for 0.707 of maximum positive signal peaks

I_3 = anode current for 0.707 of maximum negative signal peaks

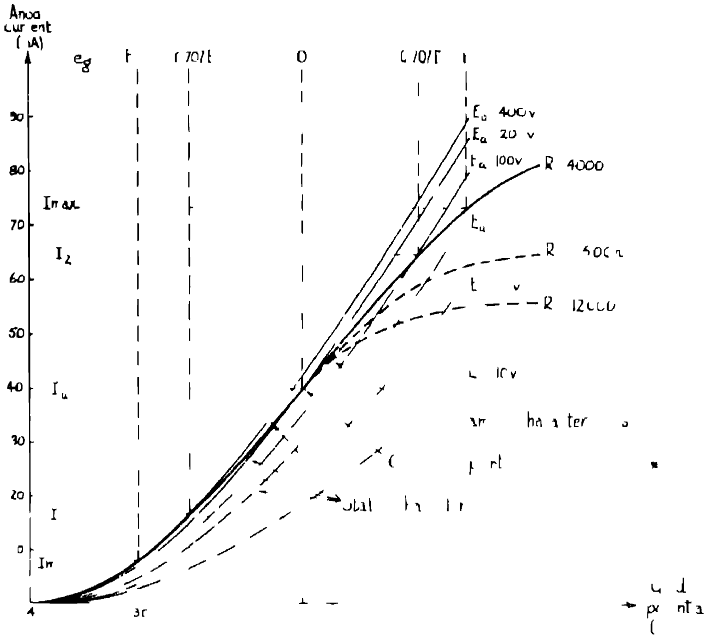


FIG. 389.—Dynamic mutual characteristics of a pentode, showing notation for calculation of harmonic distortion.

With this notation it can be shown that —

$$\text{Direct current component} = A_0 \simeq \frac{\frac{1}{2}(I_{max} + I_{min}) + (I_2 + I_3) + I_a}{4} \quad (33)$$

$$\text{Fundamental} = A_1 \simeq \frac{(I_{max} - I_{min}) + \sqrt{2}(I_2 - I_3)}{4} \quad (34)$$

$$\text{Second harmonic} = A_2 \simeq \frac{(I_{max} + I_{min}) - 2I_a}{4} \quad (35)$$

$$\text{Third harmonic} = A_3 \simeq \frac{(I_{max} - I_{min}) - \sqrt{2}(I_2 - I_3)}{4} \quad (36)$$

$$\text{Fourth harmonic} = A_4 \simeq \frac{\frac{1}{2}(I_{max} + I_{min}) - (I_2 + I_3) + I_a}{4} \quad (37)$$

$$\text{Total harmonic distortion} = \sqrt{A_2^2 + A_3^2 + A_4^2} \quad (38)$$

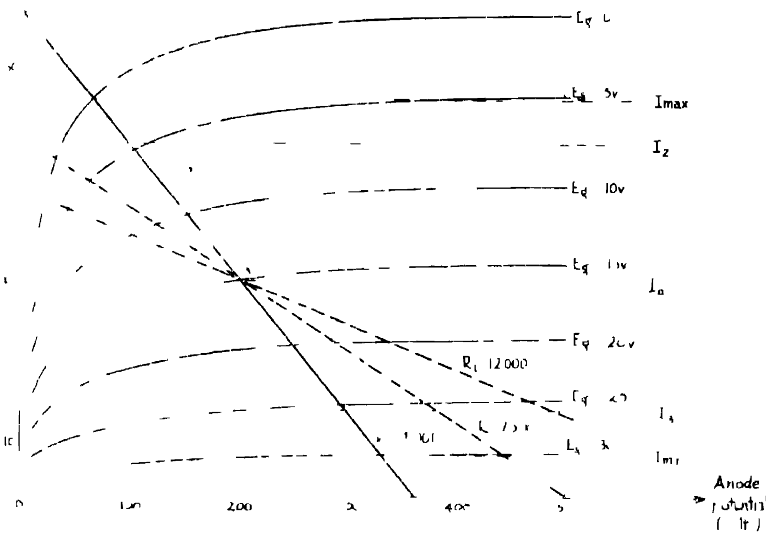


FIG. 390.—Load line for a pentode showing notation for calculation of harmonic distortion

Fig 390 shows how the various anode currents necessary for calculating distortion are derived from the load line on the anode characteristics.

Push-pull Class A amplifiers

In the push-pull amplifier two valves are arranged as shown in Fig 391.

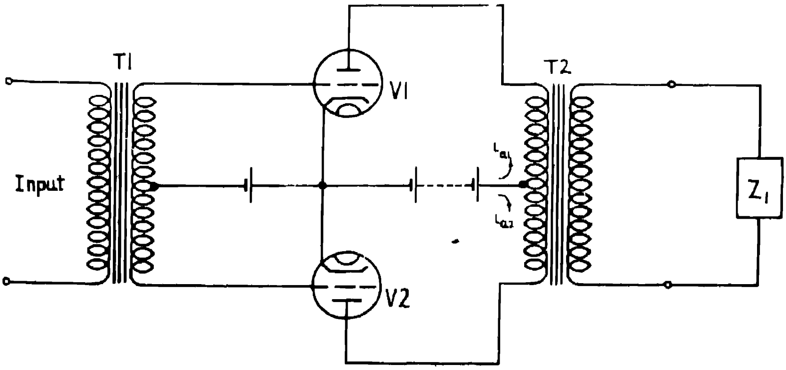


FIG. 391.—Push-pull amplifier.

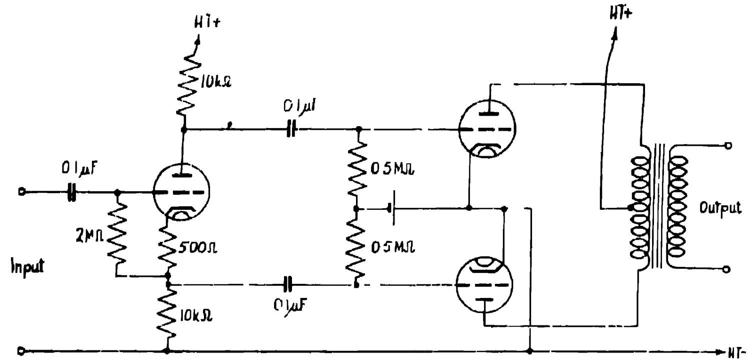


FIG. 392.—Push-pull amplifier with phase-splitter valve.

The signals applied to the two grids are 180° out of phase but equal in amplitude. These signals may be obtained by using either a centre-tapped transformer, or a phase-splitter valve (*see* Fig. 392), and the outputs of the two valves are combined by means of an output transformer having a centre tapped primary.

Fig. 393*a* shows the output waveforms of the two valves and these are combined in the load impedance as shown in Fig. 393*b*.

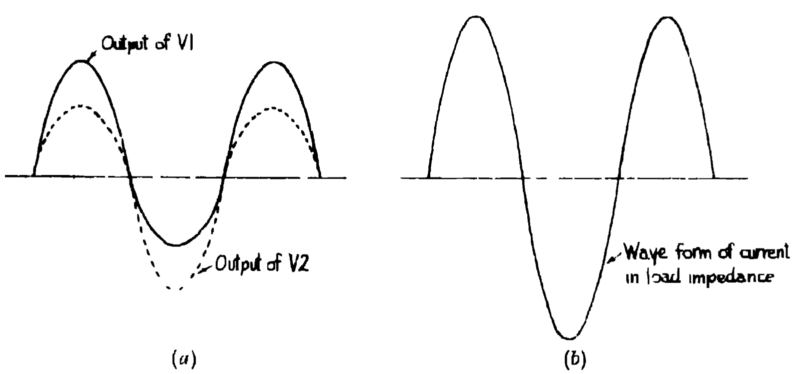


FIG. 393. Combination of outputs of two valves in class A push-pull.

In Fig. 393*a* the individual outputs of the two valves have been drawn very rich in second and other even-order harmonics, but it will be seen that the combined output is entirely free from second harmonic distortion.

Let the alternating component of the anode current flowing through the valve V_1 be given by the series:

$$i_{a1} = A \cdot e_g + B \cdot e_g^2 + C \cdot e_g^3 + D \cdot e_g^4 + \dots$$

where A, B, C, D, *etc.*, are constants, and e_g is the applied grid signal.

Provided that V_1 and V_2 have similar characteristics, the alternating component of anode current flowing through valve V_2 will then be given by —

$$\begin{aligned} i_{a2} &= A(-e_g) + B(-e_g)^2 + C(-e_g)^3 + D(-e_g)^4 + \ldots \\ &= -A\,e_g + B\,e_g^2 - C\,e_g^3 + D\,e_g^4 - \ldots \end{aligned}$$

The alternating flux produced in the output transformer will be proportional to the difference between these two currents —

$$\begin{aligned} \Phi &= k\,(i_{a1} - i_{a2}) \\ &= 2k\,(A\,e_g + C\,e_g^3 + \ldots) \end{aligned}$$

Even harmonic distortion terms thus cancel out, leaving only the odd harmonic terms.

In class A push-pull, the valves are operated in substantially the same way as for a single valve in class A, that is the operating point is at the centre of the straight portion of the dynamic characteristic and the maximum signal that may be applied to each grid is limited to the straight portion of the characteristics. The maximum signal that may be applied to the two valves in class A push-pull, and hence the maximum power output obtainable for a given percentage distortion is greater than twice that for a single valve owing to this cancellation of even harmonic distortion.

The advantages of the push-pull connection, assuming identical valves, are —

(*a*) The direct currents in the two halves of the output transformer produce opposing fluxes so that there will be no direct current saturation in the output transformer.

(*b*) The signal-frequency components of current in the HT supply cancel out and hence there is no common-impedance coupling with other stages using the same power supply.

(*c*) If the HT supply is derived from AC mains, there is no tendency for mains hum to be introduced in this stage.

(*d*) Due to cancellation of even harmonic distortion, a greater power output per valve can be obtained before the permissible distortion limit is reached.

Class B power amplifiers

In a class B amplifier, the grid bias is adjusted to "projected cut-off"—*i e* to the point P (Fig 394), where the straight portion of the dynamic characteristic when extended meets the axis of zero anode current. Except for very small input voltages, the anode current on positive half cycles is directly proportional to the input voltage, while on negative half cycles it is virtually zero. When no signal is applied the anode current is very small, and cathode grid bias is therefore unsatisfactory for this method of working.

The output (anode) current of a single valve working under class B conditions is thus seen to be a succession of pulses that are almost identical in waveform to the positive half-cycles of the input voltage, so that if a sinusoidal voltage be applied to the grid, the anode current will consist of a series of half sine waves similar

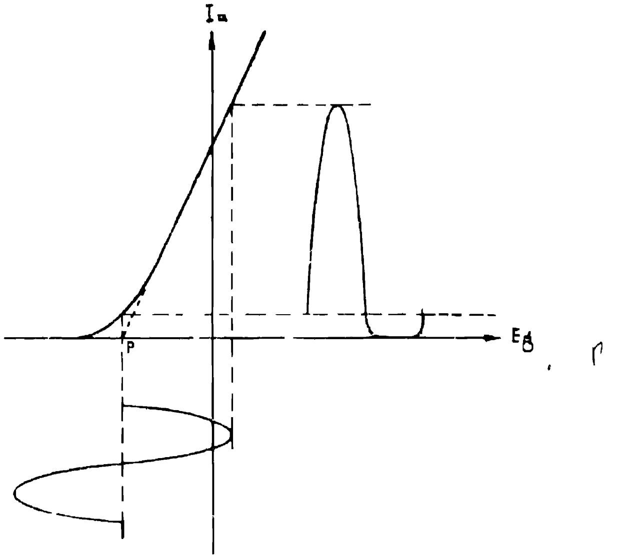
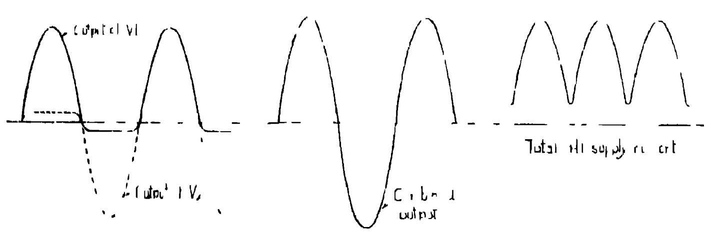


FIG. 394.—Dynamic mutual characteristic, showing class B bias conditions

to the output from a half-wave rectifier (*see* Fig. 395*a*). Current thus flows for approximately 180° of the input cycle, and the valve may be said to operate with an " angle of flow " of 180° (as opposed to 360° for a valve in class A). The distortion produced by such an arrangement is so great as to prohibit its use in an audio-frequency amplifier. If two such valves be operated in push pull, however,



(*a*) (*b*) (*c*)

FIG. 395.—Combination of outputs of two valves in class B push-pull

then each will contribute one half-cycle towards a combined anode current that is identical in waveform to the grid input (excitation) voltage; two valves in push-pull under class B conditions may thus be adjusted to give a distortionless output.

Provided that the impedance of the grid circuit is low, the dynamic characteristic of a valve is still substantially straight for small positive grid voltages, and it is possible to run into the positive-grid region on peaks without introducing excessive distortion. This will give an increase in the maximum permissible

input voltage and consequently in the output power, but in order to obtain the low grid-circuit impedance, it may be necessary to use a step-down (instead of a step-up) transformer to the push-pull stage Grid current will flow on the positive peaks of the incoming signal, and this represents a loss of power This power must be supplied by the previous stage, known as the "driver" stage, which must be designed accordingly

Thus to obtain maximum efficiency in class B operation, the grid is allowed to become positive on the peaks of input signal The maximum excitation amplitude is then over twice that permissible for the same valve when working under class A conditions, and the power output from the two valves in class B push pull may be from five to six times that obtainable from a single valve working class A Owing to the low value of anode current in the no-signal condition the efficiency of a class B amplifier is much higher than that of a class A, the theoretical maximum being 78 5 per cent , in practice efficiencies of from 50 to 65 per cent are usual

It will be seen from Fig 395c that the total HT supply current to the two valves varies at twice the signal frequency , adequate decoupling must therefore be provided, as in the case of a single valve in class A if feedback to earlier stages is to be avoided Furthermore the mean value of the anode current of each individual valve—and therefore that of the total HT supply current —rises as the input signal voltage increases , good regulation of the power supply is therefore essential A further disadvantage of class B amplifiers is that if the bias voltage is incorrect or if the two valves are not perfectly matched, severe distortion may be introduced

Class AB power amplifiers

Class AB amplifiers are used to obtain efficiencies greater than are obtainable from class A amplifiers while at the same time avoiding the critical adjustments necessary for distortion-free operation of class B amplifiers Two valves are used in push-pull and the bias is greater than that for class A operation but not as great as for class B , the operating point P is thus anywhere between the centre Q of the "straight" portion of the characteristic (*see* Fig 396) and the point B corresponding to projected cut-off bias

Where an alternating voltage is applied to the grid of a single valve operating under these conditions, the output is badly distorted, as shown in Fig 396 , when two valves are used in push-pull, however, the distortion is very small This may be verified, by drawing the individual dynamic characteristics of the two valves as in Fig 397, with the two operating points P_1, and P_2 in line From these may be drawn the combined dynamic characteristic, which can be seen to approximate to a straight line over most of its length With careful design the distortion can be kept within reasonable limits.

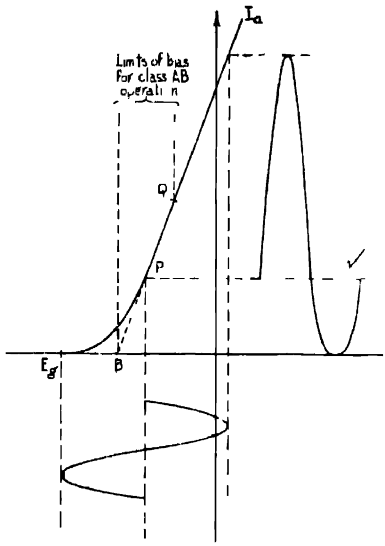


FIG 396.—Dynamic mutual characteristic, showing class AB bias conditions.

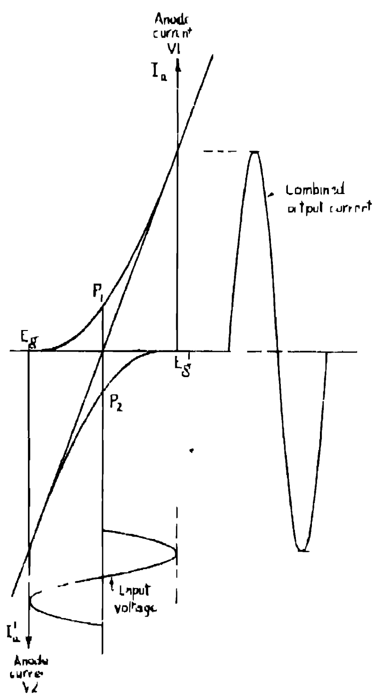


FIG. 397 —Combination of outputs of two valves in class AB push-pull

When no grid current flows, the suffix 1 may be added to the name (class AB_1). When the grid is allowed to run positive on the peaks of the incoming signal, the suffix 2 may be added (class AB_2).

As in the case of the class B amplifier, the mean value of the anode current of each valve rises as the input voltage increases, so that good regulation of the power supply is necessary. The alternating component of the total HT supply current, however, is smaller than in the case of a class B amplifier, particularly at low input signal levels, since the current taken by one valve does, to a certain extent, decrease as that taken by the other increases. Class AB amplifiers do not require quite such elaborate and carefully designed decoupling, nor do they require such careful supervision as do class B. Quite appreciable unbalance between the two valves, and deviation from the intended bias conditions may be tolerated.

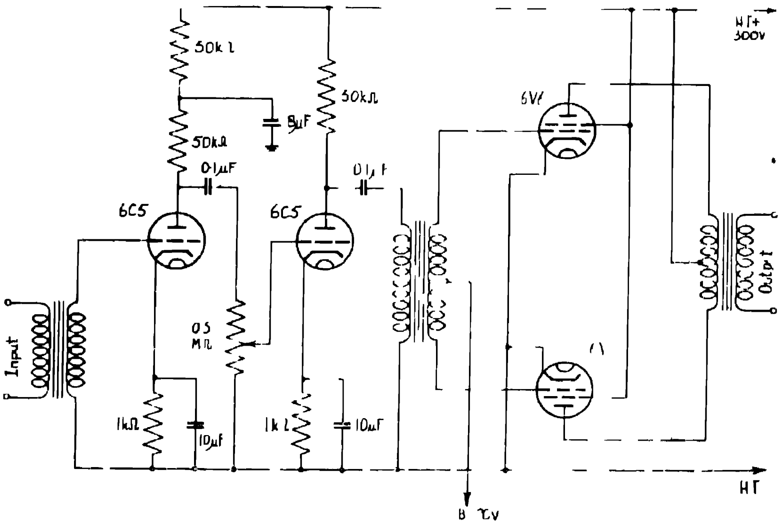


FIG 398.—Amplifier employing two 6V6's in push pull in output stage, working under class AB conditions and giving 13 watts output

The efficiency of 40 to 50 per cent obtainable in practice, together with these considerations makes the class AB amplifier very suitable for use in public address systems where medium power output and efficiency are required and moderate distortion can be tolerated.

Class C tuned amplifiers

The class C tuned amplifier differs from an ordinary tuned amplifier in that the valve is given a bias several times greater than the cut-off value. When a signal is applied, anode current flows

in pulses that last for only a fraction of a cycle (*see* Fig. 399). Owing to the distortion produced, such an amplifier is not used for audio-frequency amplification. For radio-frequency amplification, the pulses of anode current are passed through a parallel resonant circuit sharply tuned to the frequency of the input, the output across the tuned circuit is therefore sinusoidal. Since anode current

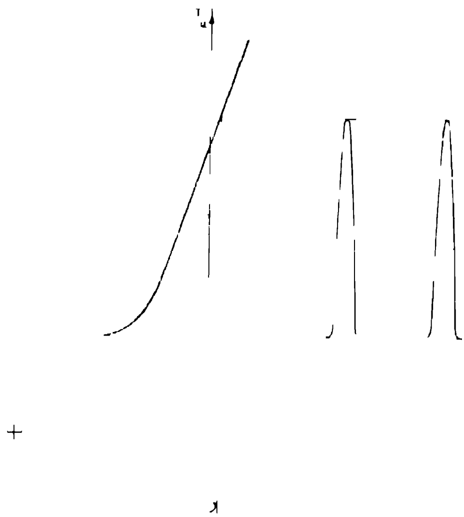


Fig. 399.—Mutual characteristic, showing class C bias conditions

flows for only a fraction of a half cycle the efficiency of a class C amplifier is very high and may approach 100 per cent. Its application, however, is limited because the output voltage is not directly proportional to the input voltage. In class C operation it is usual to allow the grid to become positive on the peaks of the input signal. The efficiency will be greater if the duration of the pulses of anode current is reduced; this also reduces the power output, although this output is developed at higher efficiency. In practice, the anode current is made to flow for about one third of a cycle, giving efficiencies of the order of 80 per cent.; this is the best compromise between high output and high efficiency.

CHAPTER 9

FEEDBACK

Soon after the discovery of the amplifying property of the triode valve, it was found that the gain of an amplifier could be increased by feeding back a portion of the output signal into the input, in such a way as to aid the incoming signal. This form of in-phase feedback was originally known as "regeneration" or "reaction", and is now known under the general title of "positive feedback". In addition to increasing the gain, it was noted that this form of feedback led to a decrease in the gain stability of the amplifier, and for this reason this method of increasing the gain of an amplifier is seldom used to-day.

On the other hand, the advantages to be obtained by sacrificing gain by the application of anti-phase or "negative" feedback have been realised only in recent years. In this form of feedback, a portion of the output signal is fed back into the input in opposition to the incoming signal. This results in a decrease in the gain of the amplifier, but, provided sufficient feedback is applied, a great improvement in the gain stability and general performance of the amplifier is obtained. In fact, it converts a valve amplifier from a device whose gain depends on numerous factors such as supply voltage and age of valves, into a precision device whose gain may be made independent of these external factors,

EFFECT OF FEEDBACK ON GAIN

Positive feedback

The application of positive feedback to an amplifier having an initial gain M will now be considered. Assume that positive

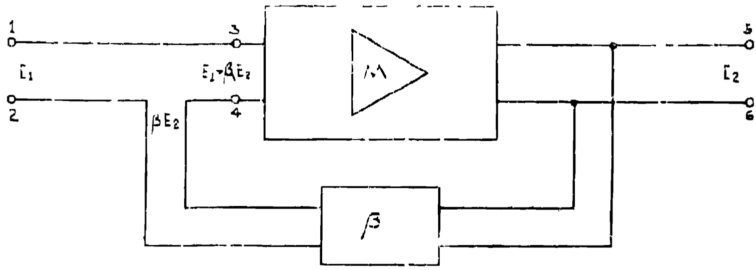


Fig. 400.—Application of positive feedback to an amplifier.

feedback is applied in such a manner, that a fraction β of the output voltage is fed back into the input circuit in phase with the incoming signal (*see* Fig. 400). In such a case, β is known as the "feedback factor".

Since a valve is a voltage-operated device as far as the grid circuit is concerned, no *power* need be fed back, and the presence of the β-network need not affect the output voltage.

Let the input voltage at terminals 1–2 be E_1, and let the output voltage at terminals 5–6 be E_2.

Since a fraction β of the output voltage E_2 is fed back into the input, the voltage fed back is $\beta \, . \, E_2$.

This fed-back voltage is in phase with the incoming signal E_1, and hence the total input at terminals 3–4 will now be $E_1 + \beta \, . \, E_2$.

The gain of the amplifier from terminals 3–4 to 5–6 is M, therefore the output voltage $E_2 = M\,(E_1 + \beta \, . \, E_2)$.

From this, it follows that :—

$$E_2\,(1 - \beta \, . \, M) = M \, . \, E_1.$$

The overall gain (M_o) of the amplifier with positive feedback is therefore :—

$$M_o = \frac{E_2}{E_1} = \frac{M}{1 - \beta \, . \, M} \qquad (1)$$

The application of positive feedback to the amplifier is thus seen to increase the gain of the amplifier from M to $\dfrac{M}{1 - \beta \, . \, M}$.

Example of positive feedback.—An amplifier has a voltage gain of 200. If $\frac{1}{400}$th of the output voltage is fed back into the input in phase with the incoming signal, find the new gain.

In this case, $M = 200$. The gain M_o with feedback is therefore :—

$$M_o = \frac{M}{1 - \beta \, . \, M}$$

$$= \frac{200}{1 - \dfrac{200}{400}} = 400$$

Thus the gain has been doubled by the application of positive feedback.

Find the new gain, if the positive feedback is so increased that $\frac{1}{250}$th of the output voltage is fed back.

The new gain M_o with feedback is :—

$$M_o = \frac{M}{1 - \beta \, . \, M}$$

$$= \frac{200}{1 - \dfrac{200}{250}} = 1000$$

Instability and oscillation

If too much feedback is applied, the amplifier may become unstable and oscillations may occur This may be considered simply as follows —

Consider the positive feedback amplifier shown in Fig 400, and assume that there is no input signal Owing to a transient effect, let a voltage e_2 appear at the output terminals 5–6 A fraction β of this voltage will be fed back *via* the feedback path into the input of the amplifier Here it will enter the amplifier as a voltage $\beta\ e_2$ and after amplification will appear in the output as $\beta\ M\ e_2$ If the gain of the amplifier is such that $\beta\ M\ e_2$ equals the initial transient voltage e_2 then the voltage will have ' regenerated ' itself and the cycle will be repeated There will thus be a continuous output voltage and the amplifier is then said to be in a state of oscillation

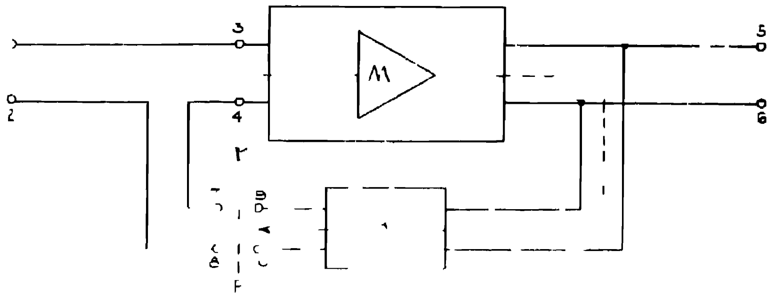


Fig 401 —Feedback loop path

It is evident that oscillations occur when $\beta\,M\,e_2 = e_2$, or in other words, when —

$$\beta\ M = 1 \tag{2}$$

" βM " is known as the " loop gain , since it represents the gain that may be measured by breaking the feedback loop at any point Considering the point P in Fig 401 it will represent the gain measured from terminals 7–8 to 9–10

It follows from the above, that an amplifier with positive feedback will be stable provided that the loop gain βM is less than one Since the gain M_o with positive feedback is —

$$M_o = \frac{M}{1 - \beta M}$$

it will be seen that when $\beta M = 1$, the gain of the amplifier is infinite It is under these conditions that an output voltage may be obtained even with no input signal.

To be more accurate, oscillations occur when $\beta M = 1, \underline{/0^\circ}$ The angle 0° is introduced here to ensure that the feedback is truly positive. M, which denotes the gain of the amplifier without

feedback, will, in general, be a vector quantity having both modulus and angle, since the output voltage will, in all probability, not only exceed the input voltage, but will also be out of phase with it. This may be represented mathematically by writing :—

$$M = |M| \angle \theta \qquad (3)$$

where $|M|$ gives the absolute ratio of the output voltage to input voltage, and θ is the phase angle between them.

The condition for oscillation is therefore that :—

$$\beta \; |M| = 1$$

and $\angle \theta = 0°$

It is only at the frequency for which *both* these conditions are fulfilled that the system will oscillate.

It should be noted that, in most practical cases, oscillation will occur if βM appears to exceed 1, provided that the angle of βM is exactly 0°. This is because in a valve amplifier $|M|$ will automatically adjust itself to give the correct oscillatory condition. There is generally no such self-adjustment as far as θ is concerned (*see* "Resistance-capacity oscillators", in Chapter 10).

Measurement of amplifier gain

This principle provides an easy method of carrying out a rough check on the gain of an audio-frequency amplifier without the use

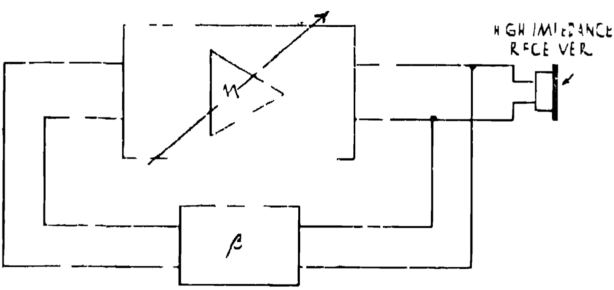


FIG. 102 Method of checking gain of amplifier

of elaborate test equipment. The output terminals are connected back to the input terminals *via* a network of known attenuation, in such a manner as to apply positive feedback. A high impedance telephone receiver is placed across the output terminals, so that any oscillation may immediately be detected (*see* Fig. 402). Then either the gain of the amplifier or the attenuation of the network is adjusted until oscillations just start. At this setting the gain of the amplifier and the attenuation of the network must be such that $\beta \; |M| = 1$. From the known loss of the network, the voltage gain of the amplifier can be calculated. Using the decibel notation, if β is, say, a 27 db loss, then $|M|$ must be a 27 db gain.

The frequency of oscillation is that frequency for which the phase shift through the amplifier and network is zero, *i.e.* the frequency for which $\beta \, . \, M$ has an angle of 0°.

Negative feedback

The case of an amplifier employing negative feedback will now be considered. In this case, the voltage fed back *opposes* the applied signal, and the voltage at terminals 3 4 (*see* Fig 403) will therefore be $(E_1 - \beta \, . \, E_2)$, instead of $(E_1 + \beta \, . \, E_2)$ as in the positive feedback case.

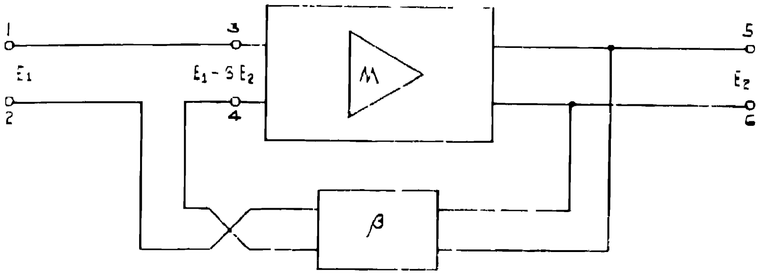


FIG. 403 — Application of negative feedback to an amplifier.

The gain of the amplifier from terminals 3–4 to 5 6 is still M; hence the output voltage $E_2 = M(E_1 - \beta E_2)$

$$\therefore \qquad E_2(1 + \beta M) = M \, . \, E_1$$

The overall gain M_o of the amplifier from terminals 1-2 to 5 6 with negative feedback is therefore :—

$$M_o = \frac{E_2}{E_1} = \frac{M}{1 + \beta M} \tag{4}$$

The application of negative feedback thus reduces the gain of the amplifier from M to $\frac{M}{1 + \beta M}$.

Example.—

Consider an amplifier having a voltage gain of 240. Let $\frac{1}{60}$th of the output voltage be fed back into the input in opposition to the incoming signal; that is, $\beta = \frac{1}{60}$. The new gain, with feedback, is :—

$$M_o = \frac{M}{1 + \beta \, . \, M} = \frac{240}{1 + \frac{240}{60}} = 48$$

The voltage gain is seen to be reduced from 240 to 48 by the application of this amount of negative feedback.

Application of a large amount of negative feedback

Considering the result that the gain with negative feedback is $\frac{M}{1 + \beta M}$, it will be seen that, if βM is large compared with 1,

then the denominator may be considered to be approximately equal to βM, and the gain M_o becomes :—

$$M_o = \frac{M}{1+\beta M} \simeq \frac{M}{\beta M} = \frac{1}{\beta} \quad \text{provided } \beta M \gg 1 \qquad (5)$$

This approximates to the practical case of applying a large amount of feedback to an amplifier having a very high inherent gain. The fact that the effective gain (M_o) of such an amplifier approximates to $\frac{1}{\beta}$ is very important, since it means that the gain is now independent of M. Any factor that may cause a change in M will not alter the effective gain of the negative feedback amplifier. Thus while changes in valves, variations in the amplification factor of a valve with age, and variations in supply voltage may have a large effect on the gain (M) of the amplifier before feedback is applied, they will nevertheless have little effect on the gain after negative feedback has been applied, since these factors do not affect β. The gain of such an amplifier is said to have been "stabilised" by the application of negative feedback.

Example.—

Consider an amplifier having a voltage gain of 20,000, and let $\frac{1}{50}$th of the output voltage be fed back into the input in opposition to the incoming signal.

This reduces the gain of the amplifier down to :—

$$M_o = \frac{20{,}000}{1 + \frac{1}{50} \cdot 20{,}000} = 49 \cdot 9$$

Suppose that, for any reason, the inherent gain of the amplifier drops to 10,000. This drastic reduction in gain reduces the gain of the amplifier with negative feedback to :—

$$M_o' = \frac{10{,}000}{1 + \frac{1}{50} \cdot 10{,}000} = 49 \cdot 75$$

Any other method of reducing the overall gain of the amplifier (*e.g.*, by an attenuator in the input) would still give a reduction by one-half when the inherent gain M dropped from 20,000 to 10,000, whereas after the application of negative feedback the overall gain drops only from 49·9 to 49·75

Adjustment of gain of a negative feedback amplifier

From the foregoing, it follows that once a large amount of negative feedback has been applied, any normal gain control that may be applied to the amplifier itself within the feedback loop will be ineffective. Alterations in the overall gain can, however, be made by making β variable. In practical cases, two methods of gain control are adopted :—

(*a*) β is fixed ; the amplifier has a constant gain, and alterations in "apparent gain" are obtained by attenuating the input signals by means of a network or potentiometer *outside* the feedback loop.

(*b*) β is variable; the gain control then takes the form of a "variable-β" control. In such a case, the gain can be varied only over a small range, otherwise too great a change in performance and in the input and output impedances of the amplifier would result.

EFFECT OF NEGATIVE FEEDBACK ON DISTORTION

In addition to the effect it has on the stability of gain, negative feedback can be used to reduce all forms of distortion produced in an amplifier.

Attenuation distortion

Attenuation distortion, which is distortion due to a variation in the gain of an amplifier with frequency, is greatly reduced by the application of negative feedback over the range of frequencies for which $\beta . M$ is large, provided that β is made independent of frequency (say, by using a simple resistive network). This follows directly from the fact that the gain in such a case approximates to $\frac{1}{\beta}$, and if β is independent of frequency, so also will be the gain.

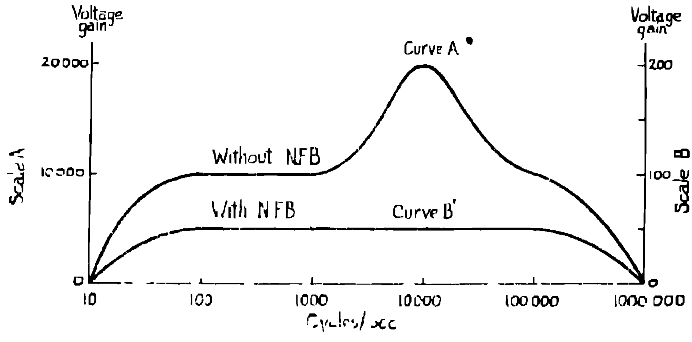


FIG 404.—Gain-frequency response curve of an amplifier with and without negative feedback

Consider an amplifier not employing negative feedback, and having a frequency response characteristic as shown by curve "A" in Fig. 404. It will be noted that such an amplifier exhibits attenuation distortion. The voltage gain rises from zero at 10 c/s to 10,000 at 100 c/s and at 1000 c/s, reaching a maximum of 20,000 at 10 kc/s, and falling to 10,000 again at 100 kc/s, dropping to zero at 1 Mc/s. From 1 kc/s to 10 kc/s, the gain rises from 10,000 to 20,000—an increase of 100 per cent.

If negative feedback is applied so that $\frac{1}{50}$th of the output voltage is fed back into the input, the gain will vary only from 49·75 at 1 kc/s to 49·9 at 10 kc/s—a variation of only 0·3 per cent. The response curve for the amplifier after feedback has been applied is shown in Fig. 404 by curve "B".

Fig. 405 shows the gain-frequency response curves of an actual three-stage amplifier taken before and after the application of negative feedback for which $\beta =$ [illegible]. It will be noted that the effects of negative feedback do not become really pronounced until the inherent gain M has become large enough to make βM very much greater than 1. This only occurs above 1000 c/s.

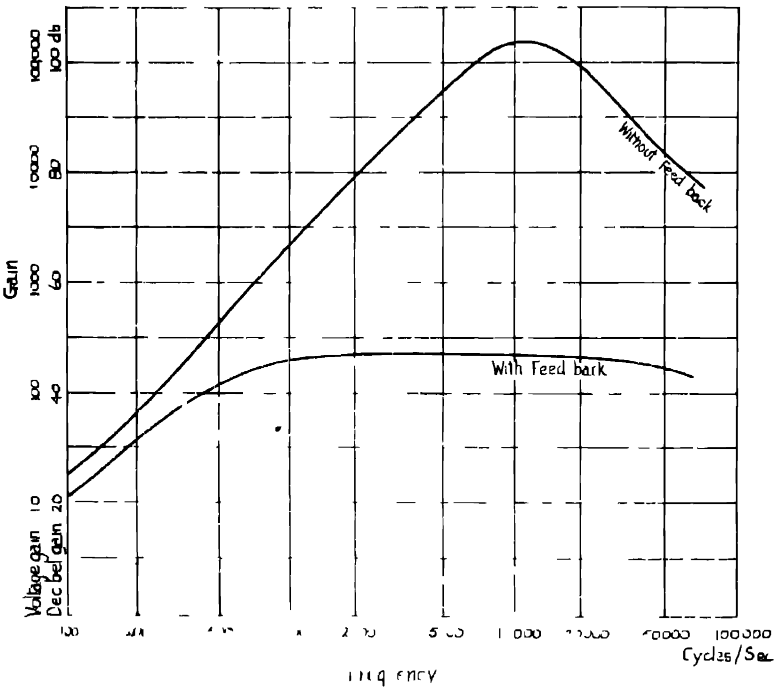


Fig. 405.—Effect of negative feedback on the gain frequency response of an actual three-stage amplifier.

So far it has been assumed that a flat frequency response curve has been desired, and this has been achieved by making β independent of frequency. It should, however, be noted that the independence of β on frequency is not an essential condition for the application of negative feedback. In fact, in certain circumstances attenuation distortion is desired and β is then made so to vary with frequency as to give the negative feedback amplifier the required gain-frequency response curve (*see* Chapter 23, " Equalisers ").

Non-linear distortion

In general the percentage of non-linear distortion in the output of an amplifier will be reduced by the application of negative feedback. This follows from the fact that, when negative feedback

is applied, any harmonics in the output will be fed back into the input, will be amplified by the amplifier, and will appear in the output 180° out of phase with those appearing in the output due to distortion of the original input signal.

However, it should be remembered that the application of negative feedback will reduce the gain of the amplifier, and hence a larger input will be required to obtain the same output with feedback than without. If this increased input has to be obtained from a pre-amplifier, care must be taken to see that this pre-amplifier is not now overloaded, otherwise the increase in distortion produced by the pre-amplifier may exceed the reduction in distortion achieved by the application of negative feedback. In the case of a multi-stage amplifier, the term " pre-amplifier " refers to all the previous stages outside the feedback loop. Thus, in a two-stage amplifier, negative feedback should never be applied to the second stage alone, unless it is certain that the first stage can supply the required increased output without overloading. It is generally better, in such a case, to apply overall feedback from the output of the second stage to the input of the first.

It can be stated that, in a multi-stage amplifier, most of the non-linear distortion will be produced by the last (output) stage, since this stage is handling the signal of largest amplitude, and that the distortion will, in general, be a function of the amplitude of the output of the stage, being very small for small outputs, but increasing as the output increases. It follows, therefore, that when considering an expression for the reduction of distortion due to the application of negative feedback, the input signal must be so adjusted that the same output is obtained in all cases, otherwise an incorrect comparison will be obtained. An approximate formula will now be deduced for the effect of negative feedback on the percentage distortion in the output of an amplifier having a non-linear gain characteristic.

Without negative feedback.—

Consider an amplifier such that, without negative feedback (Fig. 406*a*), the relationship between the input voltage E and the output voltage E_2 is given by :—

$$E_2 = M \,.\, E + N \,.\, E^2$$

$M \,.\, E$ represents the amplitude of fundamental frequencies in the output, and $N \,.\, E^2$ represents the amplitude of distortion frequencies in the output.

Then the percentage distortion in the output is :—

$$\text{Percentage distortion} = \frac{\text{amplitude of distortion frequencies}}{\text{amplitude of fundamental frequencies}} \, 100\%$$

$$= 100 \,.\, \frac{N \,.\, E^2}{M \,.\, E} \,\%$$

$$= \frac{100 \,.\, N \,.\, E}{M} \,\%$$

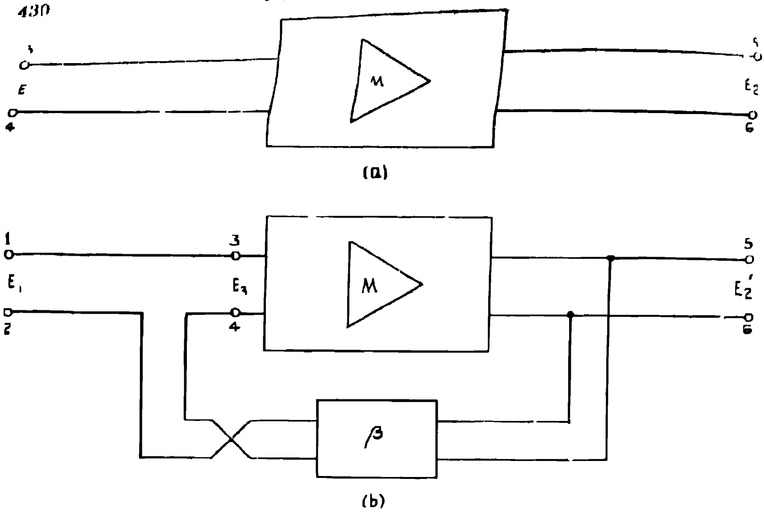


FIG. 406 Effect of negative feedback on the non linear distortion produced by an amplifier.

With negative feedback :–

Let a fraction β of the output voltage be fed back into the input in opposition to the incoming signal (Fig 406*b*) Let the input voltage at terminals 1,2 be adjusted to some value E_1 that will give the same fundamental output as before the application of NFB ; let the resulting voltage at terminals 3 4 be E_3, and let the corresponding output voltage (with feedback) at terminals 5,6 be E_2'. Then the relation between the input and output of the amplifier itself will be .

$$E_2' = M \,.\, E_3 + N \,.\, E_3^2$$

The voltage E_3 at terminals 3, 4 is given by . -

$$\begin{aligned} E_3 &= E_1 - \beta \,.\, E_2' \\ &= E_1 - \beta \,.\, M \,.\, E_3 - \beta \,.\, N \,.\, E_3^2 \end{aligned}$$

$$\therefore\ E_3 (1 + \beta \,.\, M) = E_1 - \beta \,.\, N \,.\, E_3^2$$

whence $$E_3 = \frac{E_1}{1 + \beta M} - \frac{\beta N E_3^2}{1 + \beta M}$$

Substituting in the original equation for E_2' :–

$$\begin{aligned} E_2' &= M \,.\, \left(\frac{E_1}{1 + \beta M} - \frac{\beta N E_3^2}{1 + \beta M} \right) + N \,.\, E_3^2 \\ &= \frac{M \,.\, E_1}{1 + \beta M} + \frac{N}{1 + \beta M} \,.\, E_3^2 \end{aligned}$$

The value of the input voltage E_1 must, as stated above, be such that the fundamental component of the output is the same ($M\ E$) as before the application of feedback.

Hence $$\frac{M \,.\, E_1}{1 + \beta M} = M \,.\, E$$

i.e. $$E_1 = (1 + \beta M) \,.\, E$$

whence $$E_3 = E - \frac{\beta \,.\, N \,.\, E_3^2}{1 + \beta M}$$

Then
$$E_2' = M \,.\, E + \frac{N}{1+\beta M}\cdot\left(E - \frac{\beta N E_3^2}{1+\beta M}\right)^2$$
$$= M \,.\, E + \frac{N E^2}{1+\beta M} \text{ if the distortion is small}$$

Hence the percentage distortion with NFB, for the same fundamental output ($M.E$) as before the application of NFB is :—

$$\text{percentage distortion} = \frac{100\,N\,E}{M}\cdot\frac{1}{1+\beta M}\,\%$$

The percentage distortion is therefore reduced according to the relation :—

$$\frac{\text{Percentage distortion with feedback}}{\text{Percentage distortion without feedback}} = \frac{1}{1+\beta M}$$

Thus if the gain of an amplifier is reduced by 30 db due to the application of negative feedback, the harmonic distortion present in the output will also be reduced by 30 db for the same output voltage.

It should be noted that although the gain is reduced, the design procedure for an amplifier is unaffected by the application of negative feedback. The correct operating point and anode load for, say, the output stage of an amplifier remain for all practical purposes unchanged, but by reducing the harmonic content for a given output power negative feedback will allow a still greater output power to be obtained before the permissible distortion limit is exceeded.

Phase distortion

The application of negative feedback reduces the phase shift through an amplifier. As has been stated, there will, in general, be a phase-shift through an amplifier so that the output voltage will be out of phase with the input voltage and the amplification M will be a vector quantity having both a modulus and an angle.

Let $\qquad M = |M| \angle\theta$

Then the gain M_0 with feedback is :—

$$M_0 = |M_0| \angle\theta_0 = \frac{M}{1+\beta M}$$
$$= \frac{|M| \angle\theta}{1+\beta\,|M| \angle\theta} = \frac{|M| \angle\theta}{1+\beta\,|M|\cos\theta + j\,\beta\,|M|\sin\theta}$$

Hence $\qquad \theta_0 = \theta - \tan^{-1}\dfrac{\beta\,|M|\,\sin\theta}{1+\beta\,.\,|M|\,\cos\theta}$

Thus, by the application of negative feedback, the phase-shift is reduced by an angle $\tan^{-1}\dfrac{\beta\,|M|\,.\sin\theta}{1+\beta\,.\,|M|\,.\cos\theta}$ **(7)**

Noise

The application of negative feedback reduces noise in the output of an amplifier. This noise may be due to the valves, crosstalk from neighbouring amplifiers, or power supply hum such as that produced by a poorly-smoothed HT supply. The only limitation is that the noise must have been generated or picked up by the amplifier within the feedback loop. Noise picked up by, say, an input transformer, which is outside the feedback loop, behaves as part of the incoming signal; the signal-to-noise ratio in such a case will therefore be unaffected by the application of negative feedback

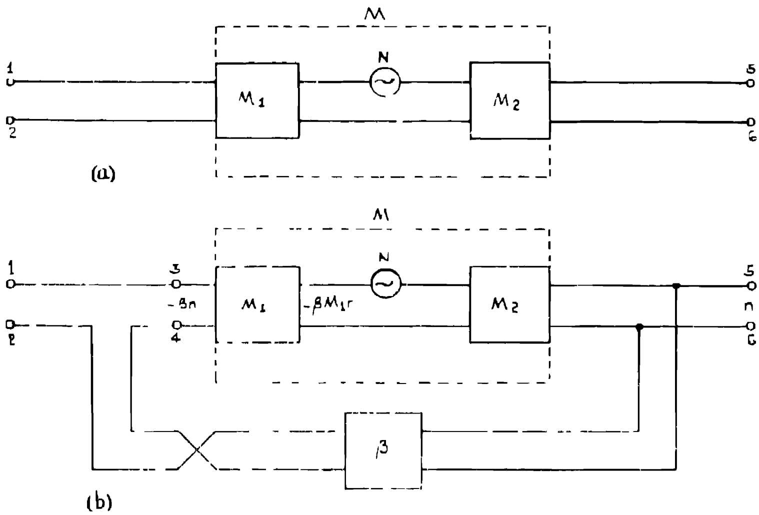


FIG. 407 Effect of negative feedback on the noise produced by an amplifier

Consider an amplifier with no feedback applied to it. It is assumed that noise present in the output is produced by a generator of EMF N situated within the amplifier (see Fig. 407*a*).

Let the gain of the amplifier preceding the generator be M_1, and let the gain after the generator be M_2. Then if M be the total gain of the amplifier without feedback,

$$M = M_1 . M_2$$

The noise N is amplified by succeeding stages of the amplifier, and will appear in the output as a voltage $M_2 . N$.

Apply negative feedback so that a fraction β of the output voltage is fed back into the input in opposition to the incoming signal, as shown in Fig. 407*b*. Let the noise voltage in the output in this case be n. A fraction β of this output is fed back, and will be applied to the input as a voltage $-\beta n$ at terminals 3–4. Thus N will be opposed by $\beta . M_1 . n$. In the equilibrium condition, the noise output voltage n is therefore given by :—

$$n = (N - \beta \, . \, M_1 n) \, . \, M_2$$

$$\therefore \quad n\,(1 + \beta \, . \, M_1 M_2) = M_2 \, . \, N$$

$$\therefore \quad n = M_2 \, . \, N \, . \, \frac{1}{1 + \beta \, . \, M_1 M_2}$$

$$\therefore \quad n = M_2 \, . \, N \, . \, \frac{1}{1 + \beta \, . \, M} \qquad (8)$$

Thus the noise with negative feedback

$$= \text{Noise without feedback} \times \frac{1}{1 + \beta \, . \, M}$$

It follows that negative feedback reduces noise in the output of an amplifier by the same factor as that by which the gain is reduced.

Example.—

$M = 240$; noise level in output without feedback $= 100$ mV.

What is the noise level when $\beta = \frac{1}{60}$?

The noise level in the output after negative feedback has been applied is :—

$$n = 100 \, . \, \frac{1}{1 + \frac{1}{60} \, . \, 240} = 20 \text{ mV.}$$

This reduction in noise level from 100 mV down to 20 mV may be compared with the reduction in gain from 240 to 48 (*see* p. 425).

Summary of effects of negative feedback on amplifier performance

Before proceeding further, the following is a summary of the effects of the application of negative feedback on the performance of an amplifier :—

(1) The gain of the amplifier is reduced.

(2) The attenuation distortion due to the variation in gain of the amplifier with frequency is reduced.

(3) The phase shift through the amplifier is reduced.

(4) Distortion due to non-linearity of the amplifier is reduced.

(5) Noise in the output of the amplifier is reduced.

(6) Change in overall amplification due to change in inherent gain (M) is reduced.

(7) Frequency response may be adjusted to any desired value by choice of a suitable feedback network.

All the above have already been mentioned ; in addition :—

(8) Both input and output impedances of the amplifier are modified. For example, the output impedance may be so modified that the voltage gain becomes independent of the output load.

The effect of negative feedback on output impedance depends on the method of applying the feedback. Typical methods will now be discussed.

VOLTAGE NEGATIVE FEEDBACK

In this type of feedback, the voltage fed back is proportional to the voltage across the anode load.

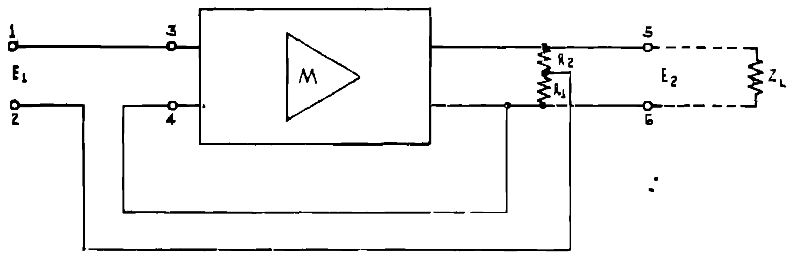


FIG. 408.—Fundamental circuit for voltage negative feedback.

The voltage fed back is usually obtained by placing a potentiometer across the anode load. Consider Fig. 408, where R_1 and R_2 constitute the potentiometer. R_1 and R_2 are large compared with Z_L; it is assumed that the input impedance to the amplifier is so large as to be considered infinite. It follows that, since $\beta = \frac{R_1}{R_1 + R_2}$, the gain M_o of the amplifier with this form of feedback will be :—

$$M_o = \frac{M}{1 + \frac{R_1}{R_1 + R_2} \cdot M} = \frac{M \cdot (R_1 + R_2)}{R_1 \cdot (1 + M) + R_2} \tag{9}$$

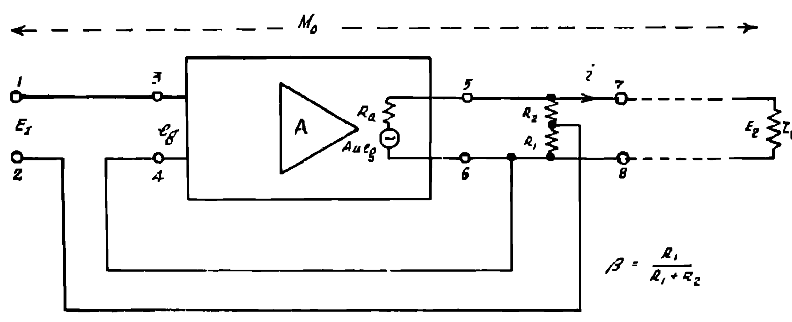


FIG. 409.—Equivalent circuit for voltage negative feedback amplifier.

An alternative approach is to consider the equivalent circuit (Fig. 409). Let the output stage of the amplifier before negative feedback is applied behave as a generator producing an EMF $A\mu e_g$, and having an internal resistance R_a. For a multi-stage amplifier employing a 1 : 1 output transformer, μ and R_a will be the amplification factor and AC resistance respectively of the last

stage, and A will be the voltage gain of all the preceding stages. In a simple single-stage amplifier not using input and output transformers, A will be unity, and μ and R_a will refer to the valve employed.

Let the current flowing in the anode load be i. Considering the circuit 6, 5, 7, 8 :—

$$A\mu e_g = i\,(R_a + Z_L)$$

$$\therefore \quad A\mu Z_L e_g = iZ_L\,(R_a + Z_L)$$

$$\therefore \quad A\mu Z_L e_g = E_2\,(R_a + Z_L) \tag{10}$$

But $\quad e_g = E_1 -$ voltage fed back

$$\therefore \quad e_g = E_1 - \beta E_2 \tag{11}$$

Substituting in (10) :—

$$A\mu Z_L\,(E_1 - \beta E_2) = E_2\,(R_a + Z_L)$$

$$\therefore \quad E_2\,(R_a + Z_L + A\mu\beta Z_L) = E_1 A\mu Z_L$$

Thus
$$M_o = \frac{E_2}{E_1} = \frac{A\mu Z_L}{R_a + Z_L\,(1 + A\mu\beta)} \tag{12}$$

Effect on output impedance

When negative feedback is applied to an amplifier an apparent change occurs in the AC resistance of the valve employed in the output stage. Since the output impedance of the amplifier depends on this AC resistance, a change in output impedance will take place when feedback is applied. The effect of voltage negative feedback on the output impedance of an amplifier will now be considered.

Without feedback.—

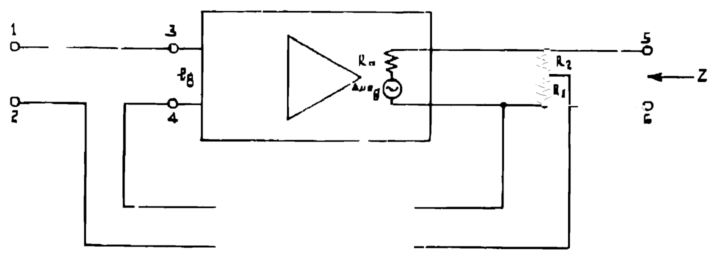


FIG. 410*a*.—Output impedance of amplifier without negative feedback.

The impedance Z looking back into the amplifier is :—

$$Z = R_a \tag{13}$$

assuming that R_1 and R_2 are large compared with R_a. If R_1 and R_2 are not large compared with R_a then the output impedance will be reduced by the presence of this shunt resistance in both this and the feedback case.

With feedback.—

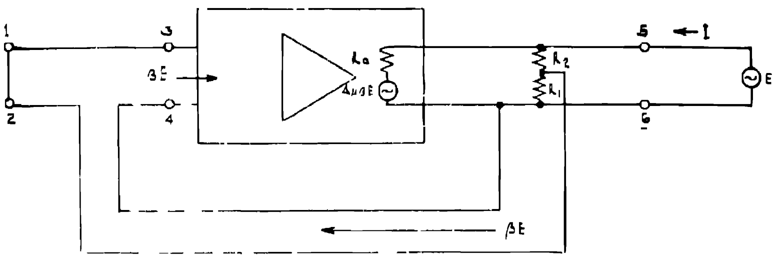


FIG. 410*b*—Output impedance of amplifier with negative feedback.

Let a generator of EMF E be applied, and let a current I flow. Then, by Ohm's Law, the new impedance looking back into the amplifier, Z_f, is :—

$$Z_f = \frac{E}{I}$$

Consider the current I flowing.

$$I = \frac{E + A\mu\beta E}{R_a}.$$

$$\therefore \quad Z_f = \frac{E}{I} = \frac{R_a}{1 + \beta A \mu}.$$

$$\therefore \quad Z_f = \frac{Z}{1 + \beta A \mu}. \tag{14}$$

Voltage negative feedback thus decreases the output impedance of the amplifier in accordance with the relation :—

$$\text{Output impedance with feedback} = \frac{\text{Output impedance without feedback}}{1 + \beta A \mu}$$

Note that while the gain is reduced by the factor $(1 + \beta . M)$, *the output impedance is reduced by the factor* $(1 + \beta A\mu)$ —NOT $(1 + \beta . M)$.

Single-stage voltage negative feedback amplifier

Fig. 411 represents a single-stage negative feedback amplifier employing voltage negative feedback. $R_1 + R_2$ will be large compared with Z_L; C is inserted to prevent HT from reaching the grid, and will have a reactance small compared with $(R_1 + R_2)$ over the working range.

If the valve has an amplification factor μ, and an AC resistance R_a, then the voltage gain M without negative feedback is :—

$$M = \frac{\mu . Z_L}{R_a + Z_L} \tag{15}$$

The voltage gain M_a after feedback has been applied is :—

$$M_a = \frac{M}{1 + \beta M}$$

i.e.
$$M_o = \frac{\mu . Z_L}{R_a + Z_L + \mu . \beta . Z_L} \tag{16}$$

It will be noted that this result can be obtained from equation 12 by putting $A = 1$.

Equation 16 may be rewritten as :—

$$M_o = \frac{\mu' . Z_L}{R_a' + Z_L} \tag{17}$$

where
$$\mu' = \frac{\mu}{(1 + \beta . \mu)}$$

and
$$R_a' = \frac{R_a}{(1 + \beta . \mu)}$$

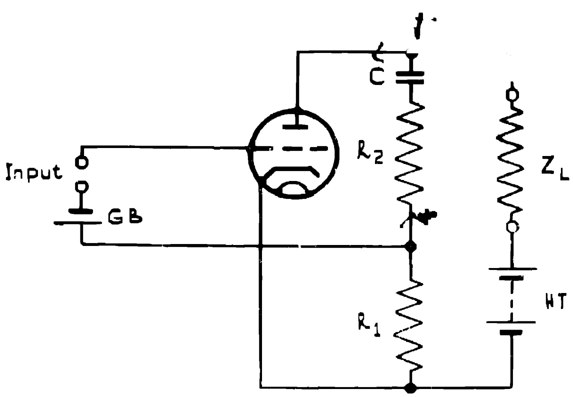


FIG. 411.— Single-stage amplifier with voltage negative feedback.

Thus, after feedback has been applied, the valve behaves as if both its amplification factor and its AC resistance had been reduced in the same ratio—*i.e.*, multiplied by $\frac{1}{(1 + \beta . \mu)}$. This reduction in impedance is of great importance in the output stage of an amplifier, since it facilitates the matching of a high-impedance valve, such as a pentode, to a low-impedance load.*

In the limiting case, when a large amount of feedback has been applied ($\beta . \mu \gg 1$), the valve impedance becomes so low that the valve behaves as a constant-voltage generator ; that is, the output voltage is independent of the anode load Z_L.

* It will be realised from this that although triode valves are used as examples throughout this chapter, the arguments apply equally well to tetrode and pentode valves.

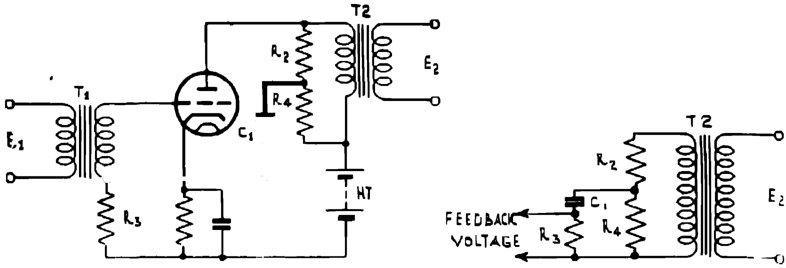


FIG. 412.—Example of single-stage amplifier with voltage negative feedback.

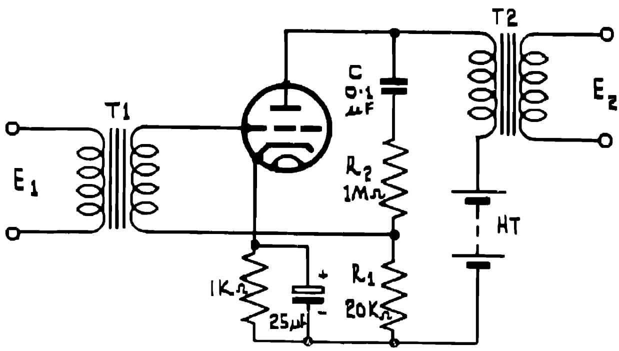


FIG. 413.—Example of single-stage amplifier with voltage negative feedback.

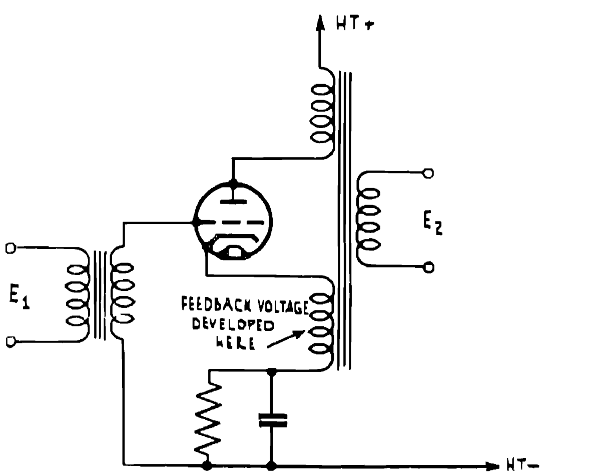


FIG. 414.—Example of single-stage amplifier with voltage negative feedback.

Let this equal $$\frac{\mu Z_L}{R_a' + Z_L} \qquad (27)$$

where $$R_a' = R_a + r(1 + \mu) \qquad (28)$$

Thus after current negative feedback has been applied, the valve behaves as if its amplification factor is unchanged, but its internal impedance has been increased by $r(1 + \mu)$.

If $\mu r \gg (R_a + Z_L)$ the gain approximates to $\frac{Z_L}{r}$. The output voltage, E_2, (which is equal to $\frac{Z_L \cdot E_1}{r}$), is proportional to Z_L. The output current, i, (which is equal to $\frac{E_2}{Z_L}$, i.e. to $\frac{E_1}{r}$), is independent of Z_L. The valve impedance thus becomes so high that the valve acts as a constant-current generator, providing an output current independent of the value of the load impedance.

Example.—

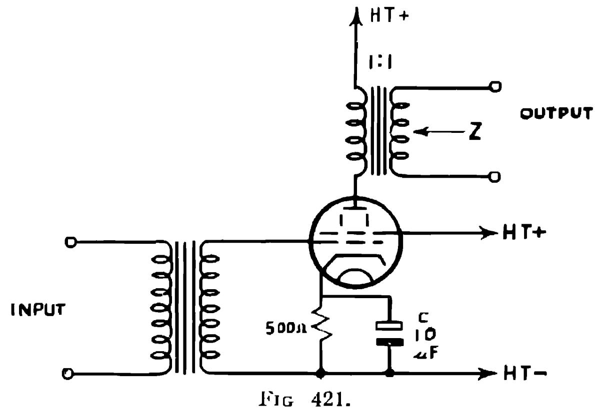


FIG. 421.

A 6V6 valve has the following constants:—

$R_a = 50{,}000$ ohms. $\qquad \mu = 200$.

It is used in a voltage amplifier (*see* Fig. 421).

Determine the effect on the output impedance Z of removing condenser C.

Without feedback:—

$$Z = R_a = 50{,}000 \text{ ohms.}$$

With feedback:—

$$\begin{aligned} Z_f &= R_a + r(\mu + 1) \\ &= 50{,}000 + 500 \times 201 \\ &= 150{,}500 \text{ ohms.} \end{aligned}$$

Removal of C thus increases the output impedance by 100,000 ohms.

Practical methods of applying current negative feedback

Single-stage current feedback

Fig 422 shows current negative feedback applied to a single amplifier stage, where the feedback resistance is (*a*) equal to, (*b*) smaller than, and (*c*) larger than, the cathode bias resistance.

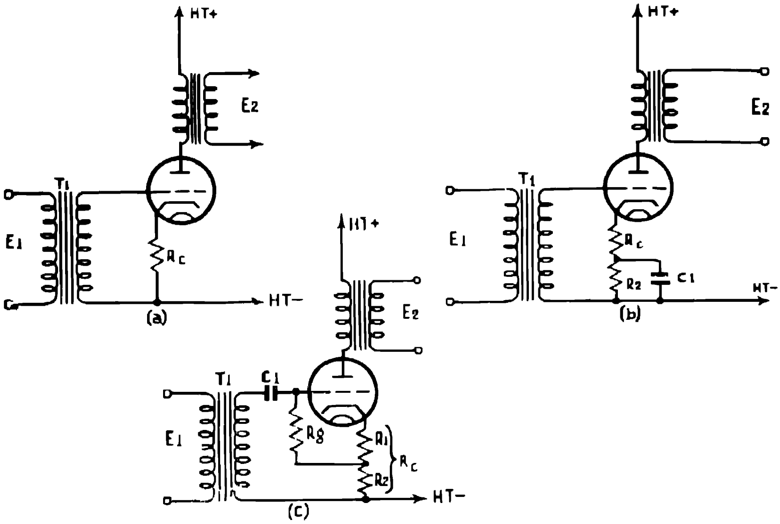


Fig. 422 Practical methods of applying current negative feedback to a single stage amplifier
(*a*) R_c provides both bias and negative feedback
(*b*) R_c provides negative feedback $R_c + R_2$ provide bias
(*c*) $R_1 + R_2 = R_c$ provides negative feedback, R_1 provides bias
($R_g \simeq 2$ megohms)

Single-stage current feedback varying with frequency

Figs 423 and 424 show how current negative feedback can be applied to make the gain increase or decrease with frequency

In Fig 423, X is comparable with R over the working range Negative feedback applied depends on frequency, being greatest at the lowest frequency The greater the negative feedback, the lower the gain.

In Fig. 424, X_L increases with increase in frequency, and therefore the negative feedback increases Thus the gain decreases with increase in frequency

Multi-stage current feedback

In Fig. 425, single-stage current negative feedback has been applied by not decoupling the cathode resistors r_1 and r_2 Feedback over two stages has been applied, since all current through the load (T_2) flows through part (R) of r_1, developing a voltage that is in opposition to the incoming signal.

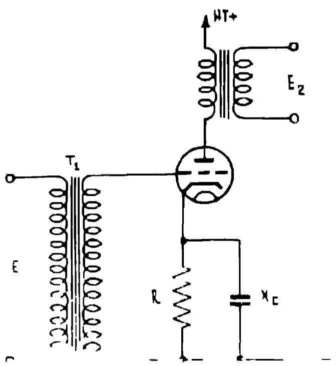


FIG 423 Single stage current feedback varying with frequency—gain increases with increase in frequency

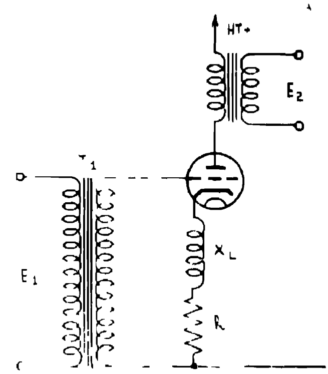


FIG 424 Single-stage current feedback varying with frequency—gain decreases with increase in frequency

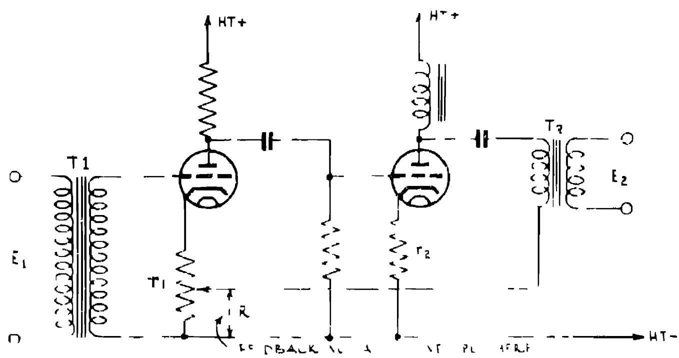


FIG 425 —Two-stage current negative feedback amplifier

COMPOSITE NEGATIVE FEEDBACK

Simultaneous application of both current and voltage negative feedback is called "composite feedback" (*see* Fig 426).

In this form of feedback, the voltage fed back in opposition to the incoming signal depends on both the voltage and the current in the anode load. The value of the total feedback factor (β) in such a case is

$$\beta = \beta_1 + \frac{r}{Z_L} \qquad (29)$$

where

$$\beta_1 = \frac{R_1}{R_1 + R_2} \qquad (30)$$

When calculating the gain of a composite negative feedback amplifier $(r + Z_L)$ is considered to be the load, but as in the current

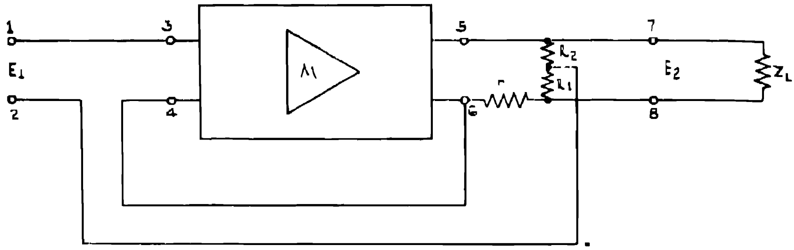


FIG. 426.—Basic composite feedback circuit.

feedback case only the voltage across Z_L must be considered when calculating the overall voltage gain.

Alternatively, it may be assumed as before, that the output stage of the amplifier behaves as a generator of EMF $A\mu e_g$ and having AC resistance R_a.

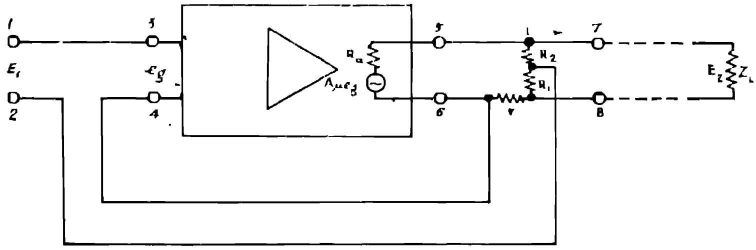


FIG. 427.—Equivalent circuit for composite feedback amplifier.

Considering Fig. 427, let the current flowing in the output circuit 6, 5, 7, 8, be i.

Then
$$e_g = E_1 - ir - \frac{R_1}{R_1 + R_2} \cdot E_2$$

$\therefore$
$$e_g = E_1 - ir - \frac{R_1}{R_1 + R_2} \cdot i\,Z_L$$

Let
$$\frac{R_1}{R_1 + R_2} = \beta_1$$

$\therefore$
$$e_g = E_1 - ir - \beta_1 i Z_L$$

But
$$i = \frac{A\mu e_g}{R_a + r + Z_L}$$

Hence
$$i = \frac{A\mu(E_1 - ir - \beta_1 i Z_L)}{R_a + r + Z_L}$$

$\therefore$
$$i\,R_a + ir + iZ_L = A\mu E_1 - A\mu ir - A\,\beta_1 \mu i Z_L$$

$$i = \frac{A\mu E_1}{R_a + r(1 + A\mu) + Z_L(1 + \beta_1 A\mu)}$$

The gain M_o with feedback is :—

$$M_o = \frac{E_2}{E_1} = \frac{iZ_L}{E_1} = \frac{A\mu Z_L}{R_a + r(1 + A\mu) + Z_L(1 + \beta_1 A\mu)} \quad (31)$$

Effect on output impedance

Consider the effect of composite negative feedback on the output impedance of an amplifier. Assume, as before, that the output stage of the amplifier behaves as a generator of EMF $A\mu e_g$ and internal impedance R_a.

Without feedback (Fig. 428).—

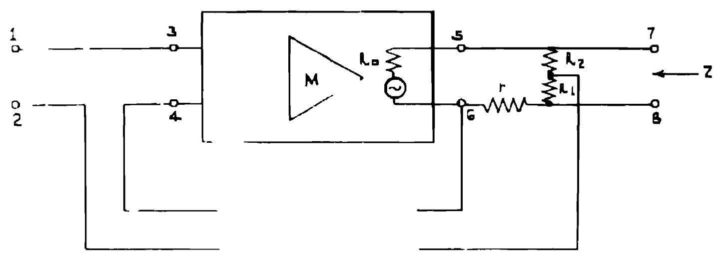


FIG. 428.—Output impedance of an amplifier without negative feedback.

The impedance Z looking back into the amplifier is :—

$$Z = R_a + r \quad (32)$$

assuming that $\qquad R_1 + R_2 \gg R_a + r$

With feedback (Fig. 429).—

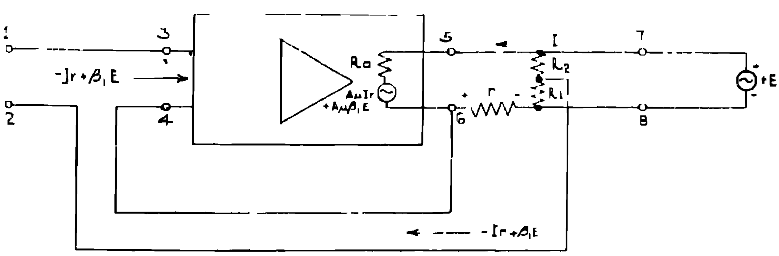


FIG. 429.—Output impedance of an amplifier with composite feedback.

Applying a generator of EMF E, let a current I flow. Hence, the new impedance Z looking back into the amplifier is :—

$$Z_f = \frac{E}{I}$$

The current I is given by :—

$$I = \frac{E - A\mu Ir + \beta_1 A\mu E}{R_a + r}$$

$$\therefore \qquad I(R_a + r + A\mu r) = E(1 + \beta_1 A\mu)$$

$$\therefore \qquad Z_f = \frac{E}{I} = \frac{R_a + r(1 + A\mu)}{1 + \beta_1 A\mu} \tag{33}$$

and

$$\frac{Z_f}{Z} = \frac{R_a + r(1 + A\mu)}{(R_a + r)(1 + \beta_1 A\mu)} \tag{34}$$

By a careful choice of β_1 and r, Z_f can be made to have any value, *i e.* less than equal to, or greater than Z

If A is large and $\beta_1 A\mu \gg 1$, Z_f approximates to $\frac{r}{\beta_1}$, which is independent of any of the valve characteristics.

Single-stage composite negative feedback amplifier

Fig. 430 represents a composite negative feedback amplifier.

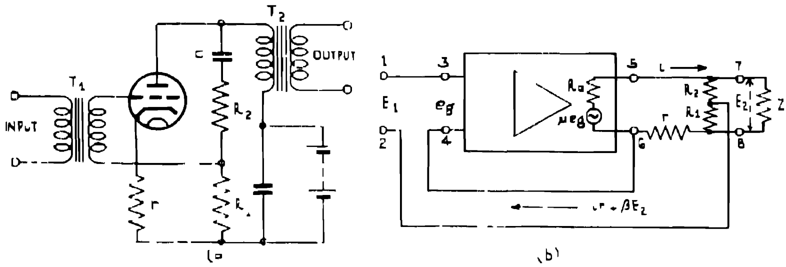


Fig. 430 – Single stage composite negative feedback amplifier

Let the valve have an amplification factor μ and an AC resistance R_a

Putting $A = 1$ in equation 31, the gain M_o with feedback is :—

$$M_o = \frac{E_2}{E_1} = \frac{\mu Z_l}{R_a + r(1 + \mu) + Z_L(1 + \beta_1\mu)}$$

$$\text{Let} \qquad M_o = \frac{\mu'' Z_L}{R_a'' + Z_L} \tag{35}$$

$$\text{where} \qquad \mu'' = \frac{\mu}{1 + \beta_1\mu} \tag{36}$$

$$\text{and} \qquad R_a'' = \frac{R_a + (\mu + 1)\, r}{1 + \beta_1\mu} \tag{37}$$

Thus after composite negative feedback has been applied the valve behaves as if its amplification factor had been reduced to $\frac{\mu}{1 + \beta_1\mu}$ and its AC resistance changed to $\frac{R_a + (\mu + 1)\, r}{1 + \beta_1\mu}$.

PLATE 23 —Two stage amplifiers employing current negative feedback

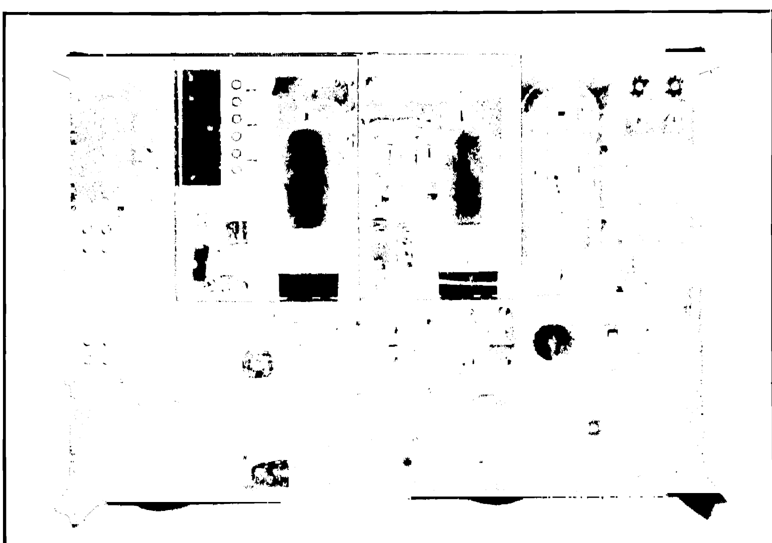


PLATE 24 —Three-stage amplifier employing composite negative feedback

Example.—

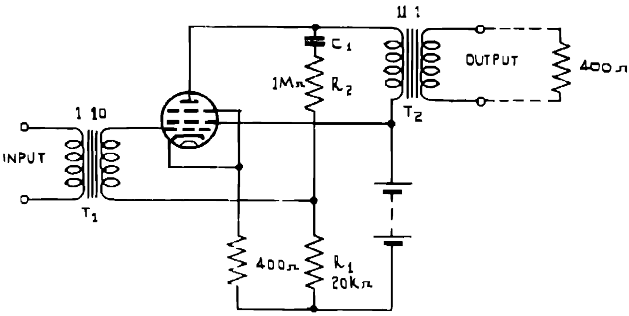


Fig. 431.

Fig. 431 shows a single-stage negative feedback amplifier employing composite feedback and designed to work into a 400-ohm load. The pentode valve used has the following constants:—

$$R_a = 10^6 \text{ ohms}$$
$$G_m = 2{\cdot}2 \text{ mA/V}$$
$$\mu = 2200$$

What is the gain and output impedance?

$$\beta_1 = \frac{R_1}{R_1 + R_2} = \frac{20{,}000}{20{,}000 + 1{,}000{,}000} = \frac{1}{51}$$

Hence $$\beta_1\mu = \frac{1}{51} \,.\, 2200 = 43{\cdot}2$$

The anode load Z_L on the valve is the reflected impedance of 400 ohms through transformer T_2.

$$\therefore \quad Z_L = 400 \times 11^2$$
$$= 48{,}400\ \Omega$$

The gain of the stage M_v with negative feedback is:—

$$M_v = \frac{\mu Z_L}{R_a + r(1 + \mu) + Z_L(1 + \beta\mu)}$$
$$= \frac{2200 \,.\, 48{,}400}{1{,}000{,}000 + 400(1 + 2200) + 48{,}400(1 + 43{\cdot}2)}$$
$$= 26{\cdot}5$$

Since the input transformer T_1 is 1 : 10 step up and the output transformer T_2 11 : 1 step down the overall voltage gain from input to output terminals $= \frac{10}{11} \times 26{\cdot}5 = 24{\cdot}1$.

The impedance Z_f looking back into the anode of the valve is:—

$$Z_f = \frac{R_a + r(1 + \mu)}{1 + \beta_1\mu}$$

$$= \frac{10^6 + 400\ (1 + 2200)}{1 + 43 \cdot 2}$$

$$= 42{,}500 \text{ ohms.}$$

The ratio of transformer T_2 is 11 : 1 step down,

$$\text{Hence output impedance} = \frac{42{,}500}{121} = 352 \text{ ohms.}$$

If feedback were not applied the output impedance would depend entirely on the R_a of the valve and on the turns ratio of T_2. It would be :—

$$\frac{10^6}{121} = 8270 \text{ ohms.}$$

SERIES AND PARALLEL FEEDBACK

Series negative feedback

In all cases of feedback so far discussed, the signal fed back has been applied in series with the incoming signal. This type of feedback is known as "series" feedback. The distinguishing

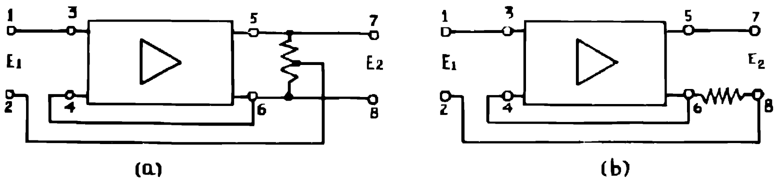


FIG. 432.—(*a*) Series voltage feedback ; input impedance increased ; output impedance reduced
(*b*) Series current feedback ; input impedance increased ; output impedance increased.

property of series feedback is that the input impedance is increased, no matter whether the feedback depends on the voltage or current in the anode load or both. This latter consideration affects only the output impedance (*see* Fig. 432*a* and *b*).

There are two standard methods of applying the feedback in series with the incoming signal. These are shown in Fig. 433*a* and *b*.

In general, method (*a*) can be adopted only when a transformer input circuit is employed. Method (*b*) is used only in conjunction with an indirectly heated valve ; single-stage current negative feedback also will automatically be applied to V_1, since no decoupling condenser can be connected across R_c.

It should be noted that a voltage applied to a resistance R in the grid circuit Fig. 433*a* will apply, between the grid and cathode of the valve, a signal that is 180° out of phase with that obtained when the voltage is applied to a resistance R_c in the cathode circuit Fig. 433*b*. Hence if the feedback obtained in (*a*) is negative, that obtained in (*b*) will be positive, and *vice versa*.

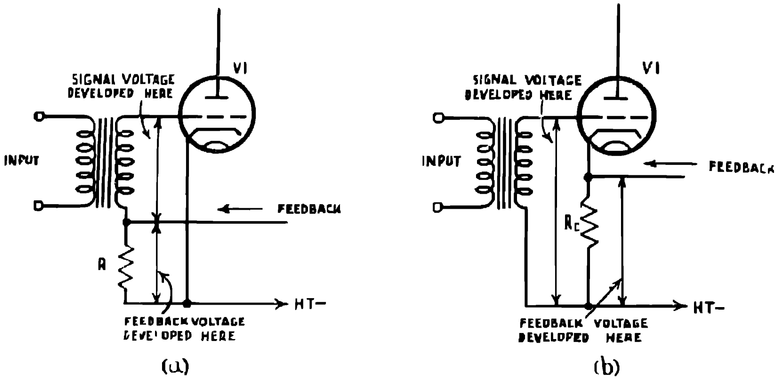


Fig. 433 Methods of applying series voltage feedback.
(*a*) Voltage developed across resistor in grid circuit.
(*b*) Voltage developed across cathode resistor.

If method (*b*) is to be applied to a pentode valve, the screen grid should be decoupled to the cathode, as in Fig. 434, and not to HT —, as is the usual practice. This is to prevent alternating components of the screen current from flowing through R_c.

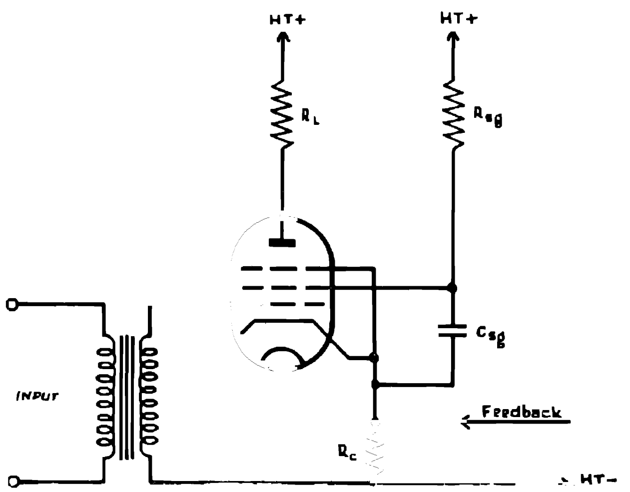


Fig. 434 —Voltage developed across cathode resistor. Note screen decoupled to cathode.

Input impedance with series feedback

To prevent either the accumulation of electrons (grid —) or the loss of electrons due to secondary emission (grid +), there must always be a conducting path between grid and cathode of resistance less than a certain value. This value depends on the particular valve (*e.g.* about 2MΩ for a high slope pentode, or 250kΩ for an

Example.—

A valve having the following constants is used in a cathode follower circuit working into a 5000-ohm load. Find the gain and output impedance.

$$R_a = 15{,}000\,\Omega$$
$$\mu = 29$$

$$\text{Gain} = \frac{\mu R}{R_a + (1+\mu)R} = \frac{29 \times 5000}{15{,}000 + 30 \times 5000} = 0{\cdot}88$$

$$\text{Output impedance} = \frac{R \,.\, R_a}{R_a + (1+\mu)R} = \frac{5000 \times 15{,}000}{15{,}000 + 30.5000} = 455\,\Omega.$$

INSTABILITY IN NEGATIVE FEEDBACK AMPLIFIERS

Consider a feedback amplifier having a gain without feedback of $M = |M| \angle\theta$. Let the feedback factor be β. So far, it has been seen that oscillations may occur if *positive* feedback is applied with $\beta|M| \geqslant 1$ and $\theta = 0°$. A negative feedback amplifier appears, at first sight, to be stable for all values of βM. However, when θ reaches $\pm\pi$ the feedback will no longer be negative but positive, since

$$\beta|M|, \angle\pm\pi = -\beta|M|, \angle 0°.$$

The change in sign of β indicates a change from negative to positive feedback. Thus if the loop gain $\beta|M|$ at the frequency or frequencies at which the loop phase-shift reaches $\pm\pi$ is equal to or exceeds 1, oscillations will, in all probability, occur. The frequency at which oscillation occurs will be that for which the absolute phase-shift round the loop is zero, the circuit adjusting itself to make βM exactly equal to 1, $\angle 0°$.

Consider a negative feedback amplifier having gain-frequency and phase-shift-frequency characteristics as in Figs. 441 and 442.

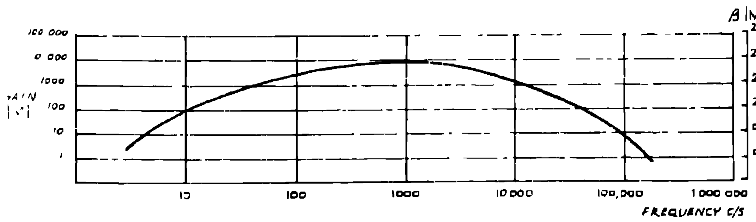


FIG. 441.—Variation of gain with frequency.

Assume that the feedback path introduces no phase-shift, and that β is independent of frequency and equal to $\frac{1}{50}$.

Let the gain of amplifier without feedback be $|M|, \angle\theta$.

At 1000 c/s the phase-shift = 0, and pure negative feedback is applied.

At 10 c/s and at 100,000 c/s, θ has reached $+\pi$ and $-\pi$ respectively. These are known as the "π" frequencies of the amplifier. At these frequencies the feedback has become purely positive feedback; but even though positive feedback is now being applied, oscillations will occur only if βM is greater than or equal to 1.

$$\text{At 10 c/s,} \quad \beta = \frac{1}{50}; \quad M = 100; \quad \beta M = 2 \text{ — unstable.}$$

$$\text{At 100,000 c/s,} \quad \beta = \frac{1}{50}; \quad M = 10; \quad \beta M = \frac{1}{5} \text{ — stable.}$$

Hence the amplifier will tend to oscillate at the lower π frequency, *i.e.* 10 c/s.

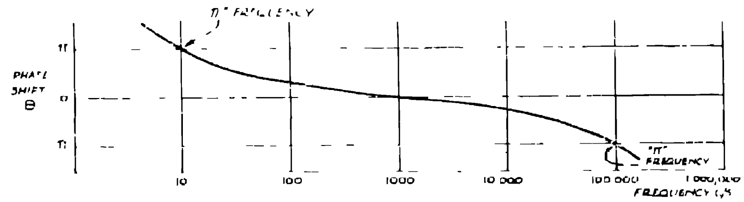


FIG. 442 —Variation of phase-shift with frequency.

To obtain full benefit from negative feedback a large amount must be applied. The chief problem in applying a large amount of feedback to a multi stage amplifier is thus one of stability. Provided that the loop phase-shift at the frequencies zero and infinity does not reach $\pm\pi$, the system can never be unstable, no matter how much negative feedback is applied; little difficulty is thus experienced over one- or two-stage amplifiers unless transformers are present in the feedback loop. With a multi-stage amplifier the loop phase-shift will probably reach $+\pi$ at some low frequency, and $-\pi$ at some high frequency. These "π" frequencies may be far outside the working range of the amplifier, but to make certain that instability does not occur it is essential to ensure that the gain $|M|$ at these frequencies is low enough to make $\beta|M|$ less than 1.

At first it seems that the simplest method of avoiding oscillation is to make certain that the amplifier gain falls off as rapidly as possible outside the working range, so that the value of M will be such as to make $\beta|M|$ less than 1 when the π frequencies are reached. Unfortunately such a drop-off in gain can be produced only by the inclusion of reactive components in the amplifier, and these components will increase the phase-shift. In fact, it can be shown that if this "cut-off" proceeds over an extended range at a rate greater than 12 db per octave (*i.e.* 4 to 1 drop in voltage gain as the frequency is doubled or halved) a phase-shift of π will be produced, and a "π" frequency reached. The design of modern feedback amplifiers is therefore concentrated on producing as rapid

as possible a fall-off in gain outside the working range, bearing this fact in mind. A decrease in gain of 10 db per octave is considered to allow ample margin for stability. Since resistance-capacity coupling in an amplifier introduces a change in gain that approaches asymptotically to 6 db/octave or a multiple of 6 db/octave, corrective networks must be used to give the required overall

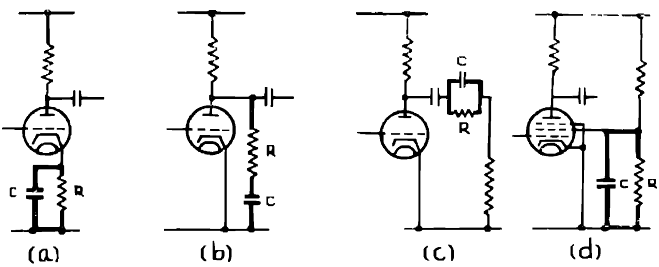


Fig. 443 Corrective networks for stabilising N.F.B. amplifiers

response. Typical networks of this type are shown in Fig. 443, where the reactance of the condenser C is comparable with the resistance R at the frequencies considered.

Conditionally stable amplifiers

It may have been noted that, when referring to positive feedback, it was stated that when $\beta|M| > 1$, oscillations *may* occur, and not that they *will*. This is because it is possible to construct an amplifier for which the feedback is apparently positive and $\beta|M| > 1$, yet the amplifier is stable; although it becomes unstable if the value of $|M|$ drops. Such an amplifier is said to be "conditionally stable".

That such an amplifier is theoretically possible may be shown from the fact that the gain M_o of any feedback amplifier is: –

$$M_o = \frac{M}{1 - \beta M}$$

where β is positive in sign for positive feedback and negative in sign for negative feedback.

To be strictly accurate, both M and β should be considered as vector quantities, since not only M, but also β may introduce phase-shift.

$$\text{Let} \quad M = |M| \angle \theta$$

$$\text{and} \quad \beta = |\beta| \angle \varphi$$

$$\text{Then} \quad M_o = \frac{|M| \angle \theta}{1 - |\beta M| \angle \theta + \varphi}$$

$$\text{and} \quad |M_o| = \frac{|M|}{|1 - \beta M|}$$

Consider the cases where the loop phase-shift is zero.

$|1 - \beta M| < 1$—*Increased gain*

$|1 - \beta M|$ being less than 1 means that β must be positive and βM must lie between 0 and 2 if the loop phase-shift is zero.

$|1 - \beta M| > 1$—*Decreased gain*

$|1 - \beta M|$ being greater than 1 means that either β can be negative and $|\beta M|$ can have any value, or β can be positive and $|\beta M|$ can be greater than 2.

In the general case where the loop gain βM will introduce a phase-shift $\angle \theta + \psi$, the gain will be decreased if $+\beta M$ plotted in a complex plane lies outside the circle of unit radius and centre $+1 + j0$, and will be increased if βM lies inside the circle.

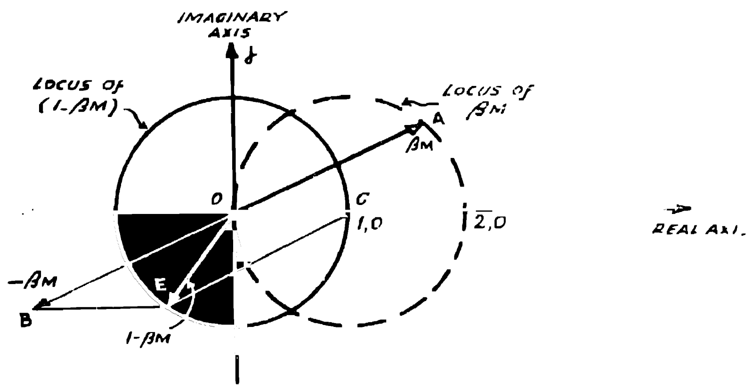


Fig. 444 Feedback circle diagram

This is shown in Fig. 444, where OA represents the vector βM, and OB represents $-\beta M$. OB is then added to the vector OC (which represents $+1$) to give the resultant vector OE, so that :—

$$\begin{aligned}\vec{OE} &= \vec{OC} + \vec{OB} \\ &= +1 + (-\beta M) \\ &= 1 - \beta M\end{aligned}$$

If $|1 - \beta M|$ is equal to 1, the point E must lie on the circle of centre O and radius 1 ; it can easily be shown that the point A must therefore lie on a circle of centre C and radius 1.

For $OCEB$ is a parallelogram,

$\therefore$ $EC = BO$ and EC is parallel to BO.

But BOA is a straight line, and $OA = OB$,

$\therefore$ $EC = OA$ and EC is parallel to OA.

$\therefore$ $OECA$ is a parallelogram.

$\therefore$ $CA = OE = 1$.

Although from this graph the effect of feedback on gain can be determined, no indication is given as to whether instability will occur.

Nyquist's condition for oscillation

Nyquist's condition for oscillation due to the application of *positive* feedback states :—

> *Let the locus of βM be plotted in a complex plane for frequencies from 0 to ∞ : then if the locus encloses the point [1, 0], oscillations will occur. If not, no matter what may be the absolute value of $|\beta M|$, oscillations will not occur.*

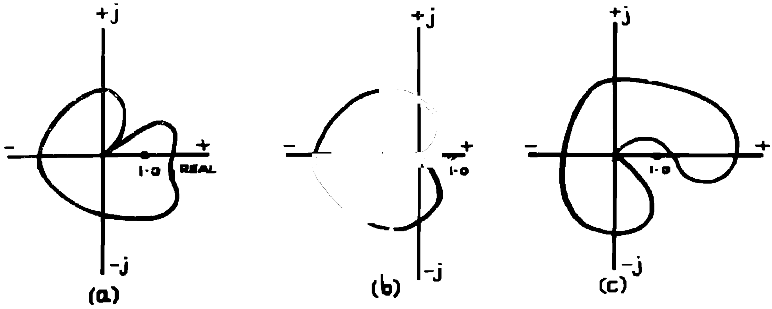


Fig. 445 Locus of βM for (*a*) unstable, (*b*) stable, and (*c*) conditionally stable amplifiers.

Hence in Fig. 445 the amplifier (*a*) is unstable, (*b*) is stable, (*c*) is conditionally stable.

Thus the only certain method of determining whether oscillations will occur, particularly in a doubtful case, is to plot the locus of $|\beta M| \angle \theta + \psi$, and to see whether the point [1, 0] is included. Alternatively, oscillations may be prevented by ensuring that $|\beta M|$ is less than 1 when $\angle \theta + \psi$ reaches $+\pi$ or $-\pi$.

APPLICATION OF NEGATIVE FEEDBACK TO LINE COMMUNICATION EQUIPMENT

Negative feedback is applied to the amplifiers employed in line communication equipment to stabilise the gain, to give the correct output impedance, to reduce distortion, and to give the required gain-frequency response.

It is essential in line equipment to prevent changes in gain due to power supply voltage variations, changes of valves, *etc.* This is very important in, for example, the case of the repeaters employed on long circuits, where a change in gain of a few db at each repeater station might result in such a large overall change in gain that the circuit would become unworkable.

With regard to obtaining the correct output impedance, two methods are adopted. The first method is to apply composite negative feedback, and this is adjusted to give not only the correct reduction in gain, but also the required output impedance. The

second method is to apply current negative feedback only. This, in itself, will give a very high output impedance, and, in order to obtain the correct impedance a shunting resistance is placed across the output. A loss of power will result, but an output impedance substantially independent of the AC resistance of the last stage will be obtained.

A particular case is that of an amplifier being used to drive a loudspeaker. Only voltage feedback is generally applied so as to provide a low output impedance to damp the resonance of the speaker.

Series (and not parallel) feedback is the form of feedback most frequently used. This tends to give the amplifiers a high input impedance and necessitates the use of an input transformer and shunting resistance to obtain the required input impedance.

In the case of carrier telephone and VF telegraph repeaters, where several bands of frequencies are amplified simultaneously by one amplifier, it is essential to reduce distortion to an absolute minimum to prevent intermodulation between the channels. Negative feedback enables this low overall distortion to be obtained in an amplifier having a high gain and high output power. It is no exaggeration to say that negative feedback has made possible the wide-band carrier systems in which several hundred channels, covering a frequency band extending up to several megacycles/sec, are all amplified simultaneously by a single amplifier without undue intermodulation.

With regard to response curves, feedback is used either with β independent of frequency to give the amplifier a flat frequency response curve over a large frequency range, or with β a function of frequency, so as to provide the required gain-frequency curve to match, say, the attenuation of the line (*see* Chapter 23, " Equalisers ").

CHAPTER 10

OSCILLATORS

Under the general heading of oscillators may be grouped all devices that provide an alternating output, whilst deriving their input from a direct current source, without changing their circuit configuration by means of switching ; this proviso excludes rotating machinery (whose commutators act as switches), and buzzers, vibrators, etc. The frequency at the output may be fixed, adjustable in steps, or continuously variable.

Oscillators fall into two main classes ; those of the first and most common class have an output waveform that is sinusoidal; those of the second class, called " relaxation " oscillators, are designed to give an output with a large harmonic content. Oscillators of the first type, which produce a sinusoidal output, can be further subdivided according to their principles of operation ; for although the majority of oscillators employ thermionic valves, they do not all use them in the same way. The only oscillator of the second class that will be considered in this chapter is the " multivibrator ". Oscillators producing saw-tooth waveforms suitable for the time-bases of cathode ray oscilloscopes will be found in Chapter 12.

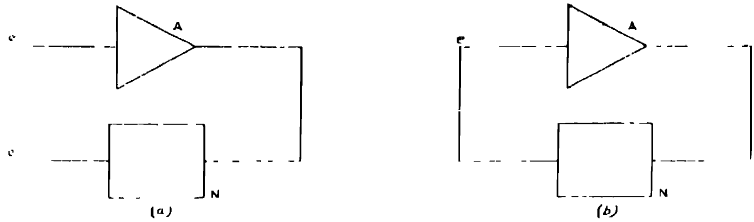


FIG. 446—Illustrating the basic principle of a feedback oscillator.

The basic principle of most valve oscillators is as follows. Fig. 446*a* shows an alternating voltage of instantaneous value *e*, and of some definite fixed frequency, applied to the input of an amplifier *A*. The amplifier output is applied to a network *N*. Now suppose that the loss of the network *N* and the gain of the amplifier *A* can be so adjusted, that the output voltage of the network is equal in *magnitude and phase* to the amplifier input voltage. If the output of the network is now connected directly to the input of the amplifier, as in Fig. 446*b*, the amplifier will continue to supply its own input, maintaining the original input voltage *e* at the original frequency.

Thus if an amplifier output, or part of the output, is fed back to the input, the system will oscillate at the frequency at which the feedback voltage is equal in magnitude and phase to the input voltage. It is evident that if an oscillator is required to oscillate at one particular frequency, it must incorporate some frequency discriminating circuit either in the amplifier itself or in the feedback network.

Suppose that the stage gain of the amplifier is M; that is, if the input voltage to the amplifier is e, the output voltage will be Me. Now suppose that the feedback network is such that the voltage fed back is a fraction β of the amplifier output voltage, that is, in the case considered, the voltage fed back to the amplifier input is βMe. The condition for maintenance of oscillation is therefore $\beta Me = e$, *i.e.* $\beta M = 1 \underline{/0^\circ}$; the angle $\underline{/0^\circ}$ is inserted as a reminder that the phase as well as the magnitude must be correct (*see* previous chapter). In a normal oscillator circuit this quantity βM is made slightly greater than unity; this, in theory, would give an oscillation of constantly increasing amplitude. In practice, however, due to the curvature of the dynamic characteristic of the valve employed the stage gain M drops slightly as the input voltage increases, and the circuit settles down to produce oscillations of magnitude such that $\beta M = 1 \underline{/0^\circ}$.

If initially βM is made much greater than unity, considerable overloading will have to occur before the stage gain drops to the value making $\beta M = 1 \underline{/0^\circ}$, and in this case the output waveform will be distorted. If a good sine-wave output is required, it is therefore important so to adjust the feedback that the circuit is only *just* oscillating.

L-C CIRCUIT OSCILLATORS

The simplest valve oscillators are those in which the output from a single-stage amplifier is fed back to its input, and in which the maintenance condition ($\beta M = 1$), and therefore the frequency of the resulting oscillations, is controlled by a resonant L-C circuit. It is important to bear in mind, that in a single-stage amplifier, a phase shift of 180° exists between grid and anode voltages. Hence if a single-stage oscillator is to be constructed, a further 180° phase-shift must be introduced into the feedback path. There are several ways of accomplishing this, the most common being the use of a transformer; other methods will be mentioned as they occur in the individual oscillator circuits.

Tuned-anode oscillator

In the tuned-anode oscillators shown in Fig. 447, a transformer is used to give the 180° phase-shift in the feedback. It is connected between anode and grid circuits, and to fix the frequency of oscillation the anode winding is tuned with a parallel condenser. The series- and parallel-fed circuits are equivalent, the advantage

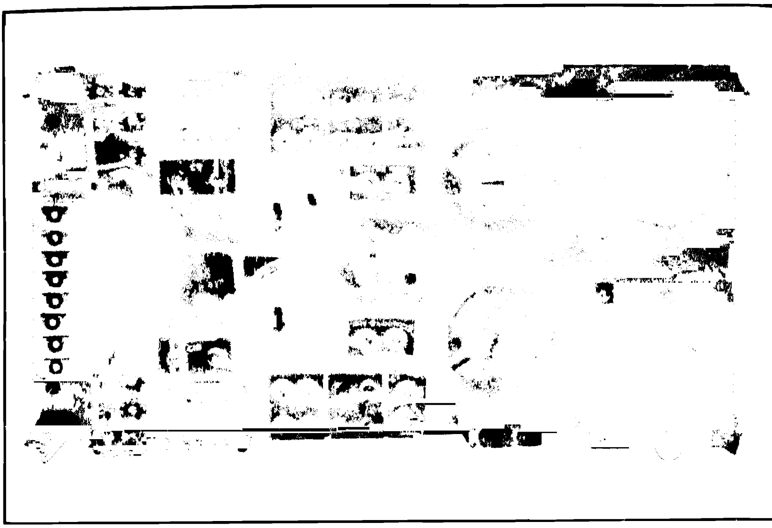
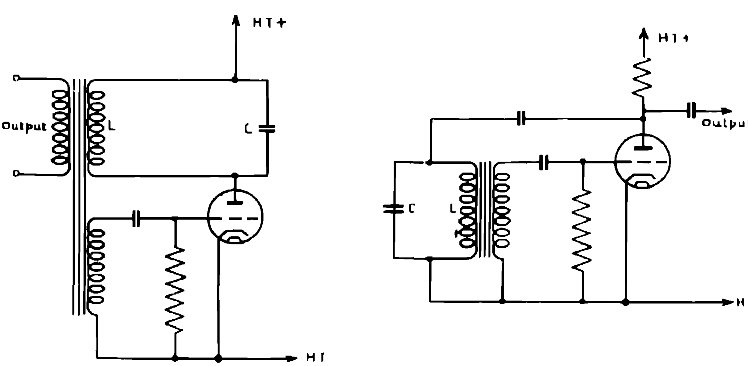


PLATE 25 —Single-stage parallel-fed tuned anode L C oscillators



(a) Series Fed (b) Parallel fed

FIG 447 —Tuned-anode oscillators

of the parallel-fed circuit being that the anode winding of the transformer does not carry the standing anode current (see transformer-coupled amplifiers, p. 393).

The frequency of oscillation is approximately the resonant frequency of L and C, *i.e.* $f \simeq \frac{1}{2\pi\sqrt{LC}}$. The condenser and resistance in the grid circuit constitute a grid leak bias circuit. The output of this, as of any other oscillator, may be taken off in a variety of ways—*e.g.* from a third winding on the transformer, *via* a condenser from the grid or anode, or from across a resistance between cathode and HT –.

Tuned-grid oscillator

Fig. 448 shows typical series- and parallel-fed tuned-grid oscillators. These are very similar to the tuned-anode oscillators already discussed, the only point of difference being that this time

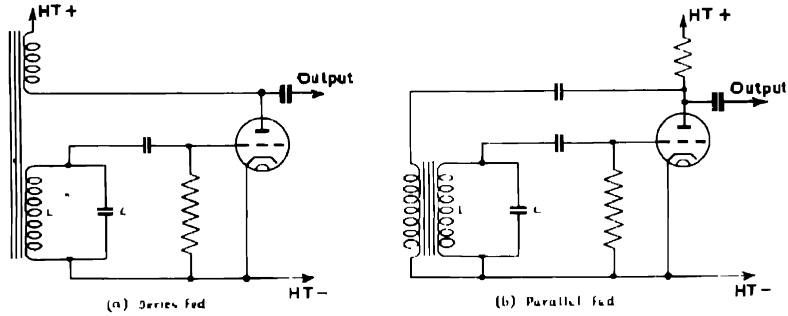


FIG. 448 Tuned grid oscillators.

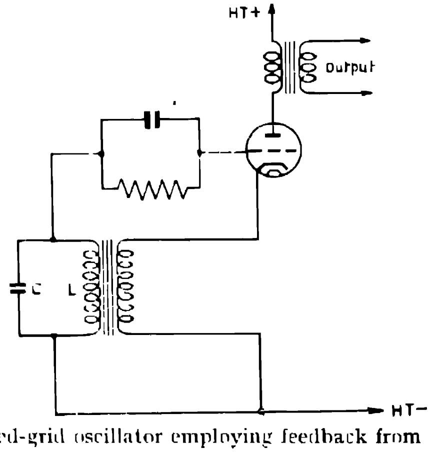


FIG. 449.—Tuned-grid oscillator employing feedback from cathode circuit

the grid winding of the coupling transformer is tuned, instead of the anode winding. The frequency of the oscillator is again approximately the resonant frequency of L and C, *i.e.* $f \simeq \frac{1}{2\pi\sqrt{LC}}$.

A modified tuned-grid oscillator is shown in Fig. 449 In this circuit, the feedback is from the cathode to the grid, instead of from the anode to the grid.

Hartley oscillator

In the Hartley oscillators the tuned circuit is connected between grid and anode, and the cathode is connected to a tapping on the inductance; in this way a 180° phase change is introduced into

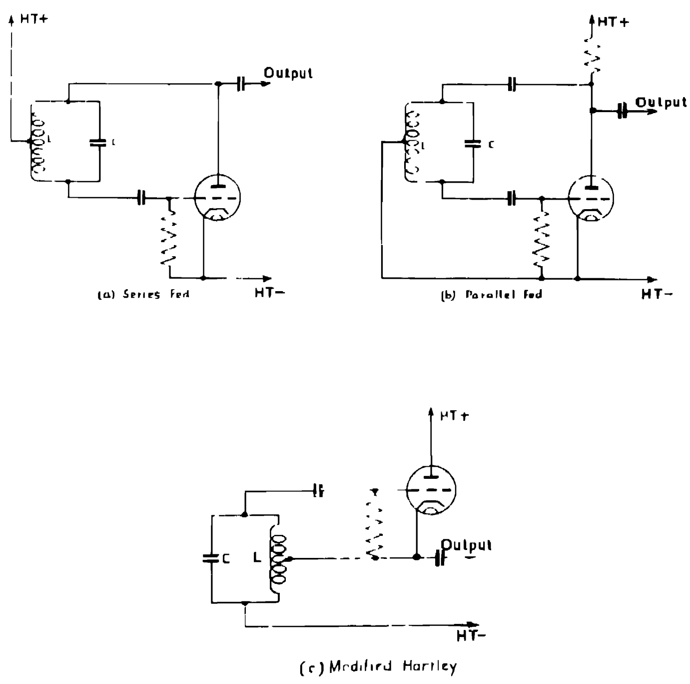


FIG. 450 Hartley oscillators.

the feedback. In the normal Hartley oscillator, either series or parallel fed, the cathode is earthed (HT −). In the modified or "inverted" Hartley oscillator, the "anode end" of the inductance is earthed. The frequency is given by the relation $f \simeq \frac{1}{2\pi\sqrt{LC}}$.

Colpitts oscillator

The Colpitts oscillator shown in Fig. 451 is similar to the shunt fed Hartley oscillator, except that the cathode is connected to an intermediate tapping on the *condenser* instead of on the *inductance*.

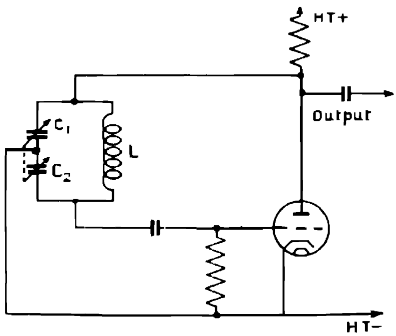


FIG. 451 —Colpitts oscillator.

The frequency is given by $f \simeq \frac{1}{2\pi\sqrt{LC}}$, where C is the capacity of C_1 and C_2 in series.

Push-pull oscillator

Fig. 452 shows a simple push-pull oscillator. The operation of the circuit is as follows. Suppose the voltage across the tuned

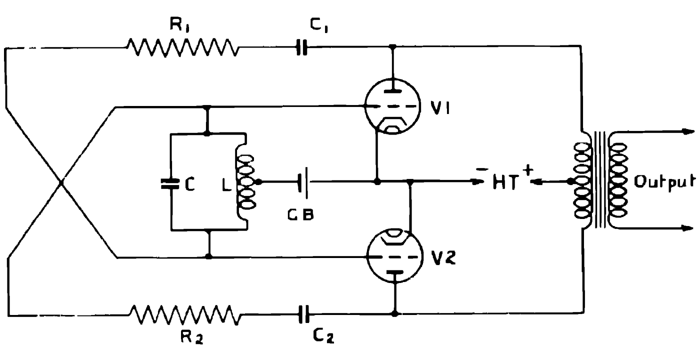


FIG. 452 Simple push-pull oscillator

circuit LC makes the grid of V_1 momentarily positive and therefore the grid of V_2 momentarily negative. This causes an increase in the anode current of V_1 and a decrease in that of V_2, producing a finite resultant current in the primary of the output transformer. The anode potential of V_1 will decrease, and that of V_2 will rise, these changes being applied to the grids of V_2 and V_1 respectively *via* the feedback circuits R_1, C_1 and R_2, C_2. Thus the feedback is

in the correct phase for the maintenance of oscillations, and the changes in the resultant current through the primary of the output transformer induce an alternating voltage in the output winding. The frequency of oscillation is approximately the resonant frequency of L and C, *i.e.* $f \simeq \frac{1}{2\pi\sqrt{LC}}$.

This push-pull connection has the same advantages as the push-pull amplifier, namely a higher output can be obtained with practically no second harmonic distortion. For this reason triode valves are used, since the distortion produced in a triode is largely second harmonic, and thus in a push-pull connection the output is practically distortionless.

Grid bias

Grid leak bias is the method normally used in oscillators, and has been employed in the circuits so far considered. It has the advantage that before oscillations begin the valve is operating at the steepest part of its characteristic, giving the best starting conditions, *i.e.* maximum stage gain. When the oscillations have reached a steady value the bias will be just sufficient to prevent the grid running too far into the region of positive grid voltage.

The time constant of the bias circuit should be greater than the periodic time at the frequency of oscillation. If the time constant is too small, excessive grid current will flow on the positive half-cycle and will cause distortion ; if on the other hand it is too large, " squegging " may occur as described below. Typical values are $0{\cdot}5$ MΩ and $0{\cdot}01$ μF for a 500 c/s oscillator ; *i.e.*, a time constant of 5 milli-seconds where the periodic time is 2 milli-seconds.

An oscillator employing this means of bias is, in general, self-starting and is capable of working satisfactorily over a wide range of values of the feedback voltage ; it is therefore the easiest type of oscillator to construct, but has two main disadvantages. The first is that if for any reason the oscillator does not oscillate, the grid is not biased back, and the anode current may rise to a sufficiently high value to cause damage to the valve ; to avoid this, an additional external bias is sometimes employed, usually in the form of cathode bias or battery bias. The second disadvantage is that the voltage output of such an oscillator is generally not constant. For this reason, where a constant output is essential, the bias is often provided externally, cathode bias or filament current bias being the usual methods employed.

It is worth noting that in the case of an oscillator employing grid leak bias, the anode current drops considerably when oscillation takes place ; this property is sometimes used to discover whether a high frequency oscillator is operating or not.

Squegging

If the time constant of the grid bias circuit is too large, regular or irregular interruptions of the oscillations may occur. This is

most likely to happen if the feedback is greater than required for in this case, initially $|\beta M|>1$, and the oscillations increase rapidly in magnitude, each positive half-cycle increasing the grid bias voltage and reducing the effective stage gain. Eventually one positive half-cycle increases the bias to such a degree that $|\beta M|<1$; if the time constant of the bias circuit is small, the bias voltage will drop during the next half-cycle to the point where $|\beta M|\geqslant 1$ and oscillations will be maintained. If, on the other hand, the time constant is large, the condition $|\beta M|<1$ will persist, and the oscillations will rapidly die out. The valve will now be biased beyond cut-off, and oscillations will not recommence

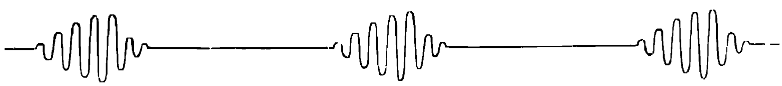


FIG. 453 Waveform of output from squegging oscillator.

until the bias condenser has discharged to the point at which anode current flows once more and $|\beta M|>1$. The waveform of the output is as shown in Fig. 453.

Mathematical treatment

The frequency of the output and the maintenance condition for any oscillator may be calculated mathematically from the equivalent circuit. This method will be demonstrated by considering a special case—that of the tuned anode oscillator shown in Fig. 454.

In the equivalent circuit, the resistance R is the resultant resistive load on the circuit, and will be made up partly by the resistance of the windings and partly by the output load, all being referred to the anode winding. Let the alternating currents through L and C be x and y respectively, so that the total alternating anode current is $x + y$. If the flow of grid current is neglected:—

$$e_g = -j\omega M x \tag{1}$$

where M is the mutual-inductance between the grid and anode windings. Now applying Kirchhoff's Law round the tuned circuit:—

$$x(R + j\omega L) + \frac{jy}{\omega C} = 0$$

i.e.
$$y = j\omega C(R + j\omega L) \cdot x \tag{2}$$

Also, applying Kirchhoff's Law round the complete circuit:—

$$x(R + j\omega L) + R_a(x + y) = -\mu e_g$$
$$= -j\mu\omega M x \text{ (from equation 1)}$$

Eliminate y using equation 2:—

$$x(R + R_a + j\omega L) + j\omega C R_a(R + j\omega L)x + j\mu\omega M x = 0$$

i.e.
$$x[R + R_a - \omega^2 LCR_a + j\omega(L + CRR_a + \mu M)] = 0$$

If the circuit is oscillating, $x \neq 0$, and therefore the condition for oscillation is:—

$$R + R_a - \omega^2 LCR_a + j\omega(L + CRR_a + \mu M) = 0$$

i.e.
$$R + R_a - \omega^2 LCR_a = 0 \tag{3}$$
and
$$L + CRR_a + \mu M = 0 \tag{4}$$

The first of these conditions gives the frequency of oscillation, namely :—

$$\omega^2 = \frac{1 + \frac{R}{R_a}}{LC}$$

i.e.
$$f = \frac{1}{2\pi\sqrt{LC}} \cdot \sqrt{1 + \frac{R}{R_a}} \tag{5}$$

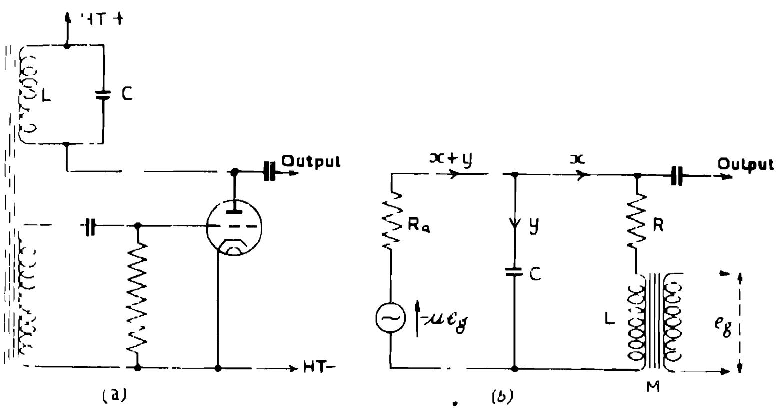


FIG. 454 Tuned anode oscillator with equivalent circuit

Note that the frequency of oscillation depends on R_a and R, *i.e.* on the output load. If, however, $R \ll R_a$ (as is usual in practice, if the oscillator is not on load)

$$f = \frac{1}{2\pi\sqrt{LC}} \text{ c/s}$$

The second condition (equation 4) may be written :—

$$M = -\frac{L + CRR_a}{\mu} \text{ henries} \tag{6}$$

and gives what is called the "maintenance condition". The fact that M takes a negative value merely shows that the transformer must be so connected as to give 180° phase-shift between anode and grid.

OSCILLATOR STABILITY

For most purposes, it is of great importance that the frequency and the output voltage of an oscillator shall be stable that is, that they should remain constant at the values to which they are adjusted. The factors on which stability depends, and means of ensuring it, will now be considered.

Frequency stability

Taking the example of the tuned-anode oscillator investigated above, the frequency is given by :—

$$f = \frac{1}{2\pi\sqrt{LC}} \cdot \sqrt{1 + \frac{R}{R_a}}$$

and it has been seen that, if $R \ll R_a$, this reduces to

$$f = \frac{1}{2\pi\sqrt{LC}}$$

In this case the frequency depends only upon the constants of the tuned circuit. Valve capacities must be included in the value of C, but these are normally constant.

Effect of temperature

It was shown in the last paragraph that the oscillator frequency was determined almost entirely by the components forming the tuned circuit. If the oscillator frequency is to remain constant for variations in temperature, it is important to ensure that the values of L and C do not vary.

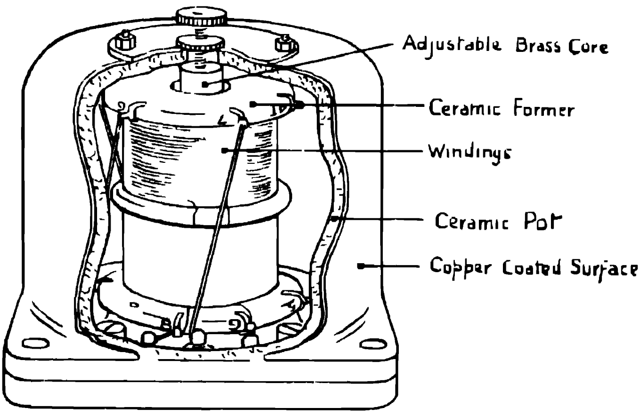


FIG. 455 Typical inductance having high temperature-stability.

With good condensers, the effect of temperature on C is very small. Condensers can be designed to have almost any desired capacity-temperature characteristic; they can be so made that the capacity will increase, decrease, or remain constant with changes in temperature. Inductances, on the other hand, are more susceptible to temperature changes, as expansion will alter the inductance. Compensation is therefore required to ensure that the inductance has a small temperature coefficient. In addition, to prevent permanent changes in inductance, the former on which the coil is wound should have the same coefficient of expansion as the wire of the coil. If screening is employed, it must be so arranged that the change in its effect on the inductance of the coil

with temperature is negligible. Fig. 455 shows an inductance wound on a ceramic former, having a low coefficient of expansion, and screened by means of a copper-sprayed ceramic pot. Such an inductance has good frequency stability.

Resistance stabilisation

Now consider the oscillator working into a load. In this case, the condition $R \ll R_a$ no longer holds, and the frequency depends on R and on R_a, that is, on the load and on the valve constants; the frequency thus varies with a change of supply voltage. This state of affairs is undesirable, and can be minimised by using parallel feed through a large resistance, which has the effect of increasing artificially the value of R_a.

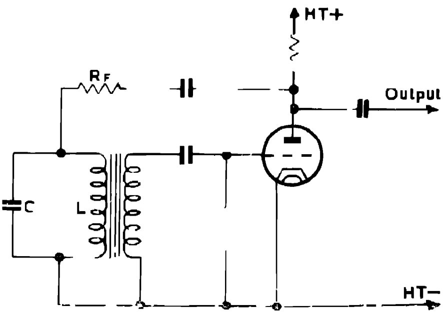


FIG. 456 Tuned anode oscillator, parallel fed, with feedback resistance to improve frequency stability

Fig. 456 shows a typical parallel fed tuned anode circuit utilising a feedback resistance R_F. This resistance has also the advantage of giving a fine control over the amount of feedback. The feedback resistance R_F must be large compared with R_a and R_L in parallel, where R_L is the load impedance in the anode circuit. Note that although variations of output load may now have little effect on frequency, they may still have a considerable effect on the maintenance condition for oscillation.

Buffer amplifier

The most satisfactory way of ensuring a high degree of frequency stability in an oscillator is, however, to use another valve as an amplifier stage following the oscillator. Such an amplifier is known as a " buffer " amplifier; and, since the oscillator works straight into the grid circuit of the buffer amplifier, the load on the oscillator is negligible and the frequency stability is high.

A typical circuit is shown in Fig. 457.

Note that in this case the tuned circuit, parallel-fed from the anode of the first valve, forms the grid circuit of the second valve.

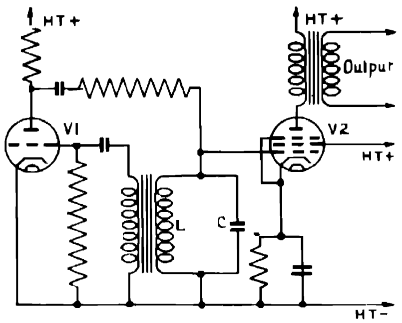


FIG 457.—Oscillator with buffer amplifier, giving high frequency stability.

Electron-coupled oscillators

An electron-coupled oscillator, an example of which is shown in Fig. 458, although employing only one valve, is equivalent to an oscillator followed by a buffer amplifier, and has the same high frequency stability.

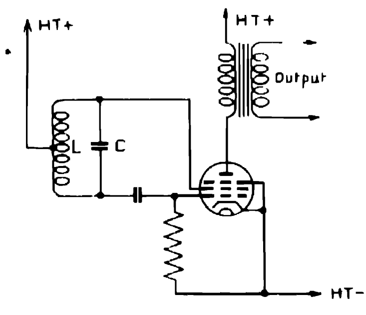


FIG. 458 – Electron-coupled Hartley oscillator.

The valve used is either a tetrode or a pentode, and the screen is used as the "anode" of the oscillator; thus in Fig. 458 the screen, the grid and the cathode are connected up in the form of a series-fed Hartley oscillator. Changes in anode load (within certain limits) have no effect on screen current and hence the load does not affect the oscillator circuit. On the other hand the oscillating voltage between cathode and grid affects the anode current; since the output is taken from the anode circuit, the valve behaves as its own buffer amplifier.

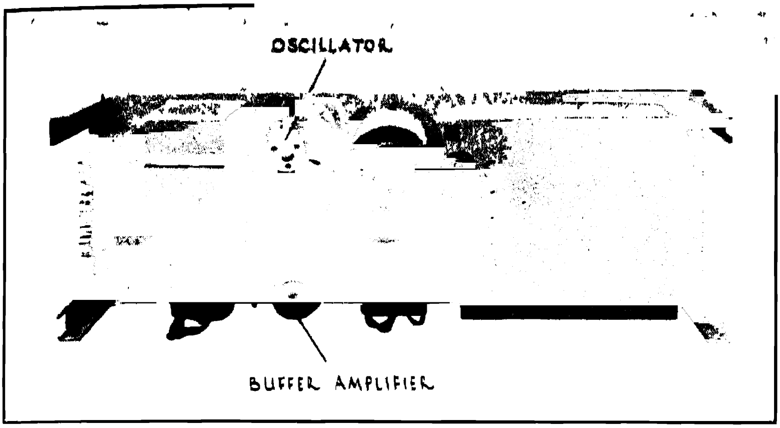


PLATE 26.—Electron coupled tuned-grid oscillator employing an additional buffer amplifier

Output voltage

There are two main factors that affect the output voltage of a valve oscillator namely, the HT supply voltage and the amount of feedback. Variations of the former produce approximately proportional variations in output voltage. The output is also increased if the amount of feedback is increased, though this will also result in increased distortion. It is never possible to forecast accurately the output voltage of an oscillator. If, however, the feedback is so adjusted that oscillation only just takes place, the output voltage will rise to the point where the valve characteristics cease to be linear, due either to cut-off at the bottom end or to saturation or grid current at the top. If *fixed* bias is used, and is so adjusted that grid current flows on a smaller grid swing than that required to cut the valve off, the peak value of alternating voltage on the grid will normally just equal the fixed bias voltage; for if the grid swing increases beyond this point, grid current will flow and the circuit will be heavily damped, thereby reducing the grid swing. It is important to remember that, although an oscillator valve may be operating under conditions that would normally produce enormous distortion, the voltage across the tuned circuit may be almost sinusoidal, due to its impedance being small at all other frequencies.

Stabilisation of output voltage

The variations of output voltage with supply voltage may be minimised by the insertion in the circuit of some form of non-linear

resistance. A tungsten filament lamp, for instance, has a resistance that rises rapidly with increase in the current through it. If such a lamp is placed in the cathode circuit without decoupling, it will provide negative feedback that increases as the current through it increases (*see* Fig. 459, which shows the carrier-frequency oscillator of a single-channel carrier telephone system).

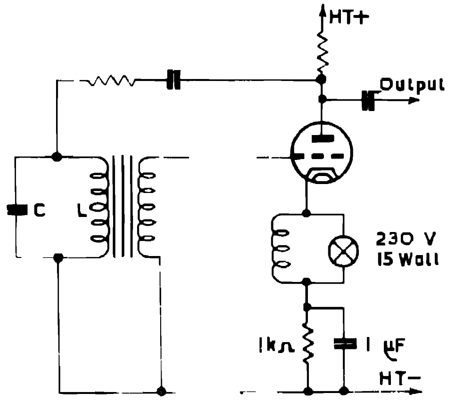


FIG. 459.—Carrier-frequency oscillator, showing tungsten-filament lamp for stabilising output voltage.

The inductance in parallel with the lamp carries most of the DC cathode current, whilst forming a negligible shunt on the lamp at the oscillator frequency. An increase in output, consequent upon an increase in HT supply voltage, therefore increases the alternating current in the lamp. This results in an increase in the lamp resistance, an increase in the amount of current negative feedback, and a reduction in the stage gain, thus stabilising the output voltage.

Neon stabiliser

Another method of stabilising the voltage output is shown in Fig. 460. This method gives better stabilisation than the last

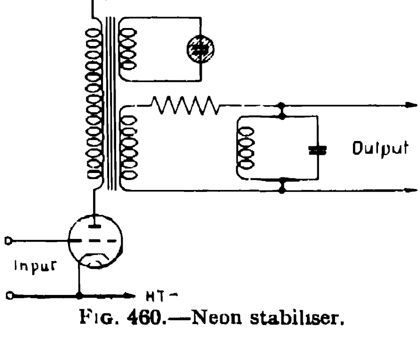


FIG. 460.—Neon stabiliser.

and is used when a stable output voltage is essential, e.g. the pilot oscillator on certain carrier telephone systems. The oscillator output obtained *via* a buffer amplifier, has a neon stabiliser connected across it. This consists of a neon tube connected across a third winding of the output transformer. The neon lamp "flashes over" and offers a low impedance when the voltage across it exceeds a certain value. This value is exceeded by the peaks of the output waveform, which are therefore cut off. The distortion introduced by this method is eliminated by a tuned circuit following the stabiliser, as shown in Fig. 460.

Lamp bridge stabilisation

Another type of oscillator makes use of the properties of a lamp bridge to obtain constant output. As explained in Chapter 5, the lamp bridge consists of a Wheatstone bridge made up with a tungsten filament lamp in one arm and fixed resistances in the

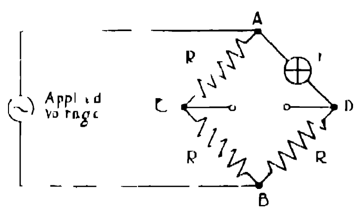


Fig. 461.—Basic circuit of lamp bridge stabiliser.

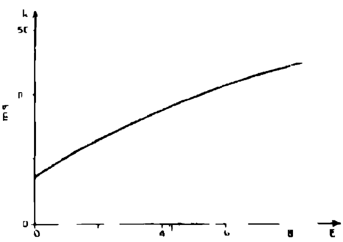


Fig. 462.—Variation of resistance with applied voltage for a 6-volt tungsten filament switchboard indicator lamp.

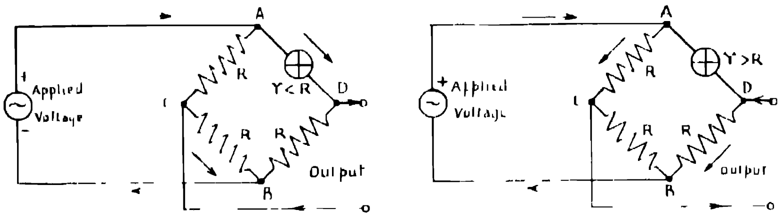
other three (see Fig. 461). The resistance of the lamp varies with the voltage across it (see Fig. 462) and hence with the voltage across the bridge.

By a suitable choice of components the bridge can be made to balance at one particular value of applied voltage, the balance being independent of frequency. Consider the positive half-cycle of an applied voltage such as to make point A positive with respect to point B. If the applied voltage is less than that required for balance, the resistance of the lamp will be smaller than the value required for balance. Since the resistance of the lamp is now less than R, point D will be at a higher potential than point C, and current will flow in the output in the direction as shown in Fig. 463*a*.

If, on the other hand, the applied voltage is greater than the value required for balance, the resistance of the lamp will be greater than R, point C will be at a higher potential than point D, and current will flow in the output in the reverse direction, as shown in Fig. 463*b*.

It will be seen that the bridge introduces a phase-shift of 180° in

the output as the applied voltage increases through the value required for balance. This effect takes place for both DC and audio-frequency voltages. The bridge is inserted in the feedback path of an oscillator in such a way as to make the feedback positive if the output voltage is below the critical value, and negative if



(*a*) Applied voltage less than that required for balance

(*b*) Applied voltage greater than that required for balance

FIG 463.—Lamp bridge, showing phase relationship between input and output on either side of balance condition

above. On switching on, the lamp is cold, and oscillations are built up since positive feedback is being applied. The oscillations will increase in amplitude until the output voltage is just below the value required to balance the bridge, and the output will be stabilised at this value. Fig. 464 gives an example of this type of oscillator.

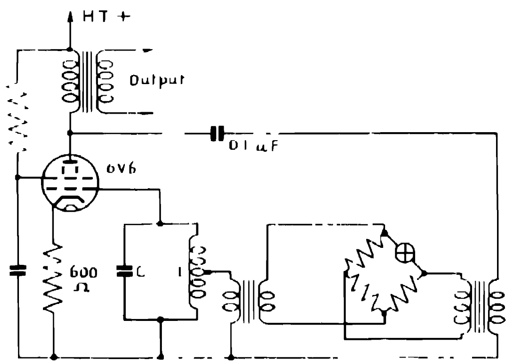


FIG. 464 Oscillator with lamp bridge stabilisation.

NEGATIVE-RESISTANCE OSCILLATORS

Consider the circuit of Fig. 465 ; a resistance r is connected across a parallel tuned circuit. Let the currents in the various parts of the network be as shown ; then considering the steady state only and applying Kirchhoff's Laws :—

$$r(x+y) - y \cdot \frac{j}{\omega C} = 0$$

$$\therefore \qquad rx = -y\left(r - \frac{j}{\omega C}\right) \tag{7}$$

and
$$r(x + y) + x(R + j\omega L) = 0$$

i.e.
$$ry = -x(r + R + j\omega L) \tag{8}$$

Multiplying equation 7 by equation 8 :—

$$r^2xy = xy\left(r - \frac{j}{\omega C}\right)(r + R + j\omega L)$$

i.e.
$$xy\left[rR + j\omega Lr - j\frac{r}{\omega C} - j\frac{R}{\omega C} + \frac{L}{C}\right] = 0$$

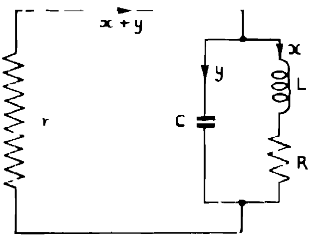


FIG. 465.—Illustrating the principle of the negative resistance oscillator.

If the circuit is oscillating, $xy \neq 0$, and the condition for oscillation is therefore :—

$$rR + \frac{L}{C} + j\left(\omega Lr - \frac{r + R}{\omega C}\right) = 0 \tag{9}$$

Thus gives the *maintenance condition* :—

$$rR + \frac{L}{C} = 0$$

i.e.
$$r = -\frac{L}{CR} \tag{10}$$

and the frequency is given by :—

$$\omega Lr - \frac{r + R}{\omega C} = 0$$

i.e.
$$\omega^2 = \frac{1 + \frac{R}{r}}{LC}$$

Hence, using equation 10 :—

$$\omega^2 = \frac{1 - \frac{R^2C}{L}}{LC}$$

i.e.
$$\omega^2 = \frac{1}{LC} - \frac{R^2}{L^2}$$

or
$$= \frac{1}{2\pi}\sqrt{\frac{1}{LC} - \frac{R^2}{L^2}} \qquad (11)$$

Therefore, if a negative resistance equal to $-\frac{L}{CR}$ is connected across a parallel tuned circuit, oscillations will be maintained at the resonant frequency of the tuned circuit

Negative resistances may be obtained in a number of ways and three methods using a thermionic valve will be discussed in some detail It is worth mentioning that a carbon arc has a negative resistance and this was the basis of the earliest radio frequency oscillators used in the ' spark ' transmitters

The dynatron oscillator

The dynatron oscillator uses a tetrode valve to provide negative resistance As has been explained in Chapter 7, the anode characteristic of a screen-grid valve has a region of negative slope

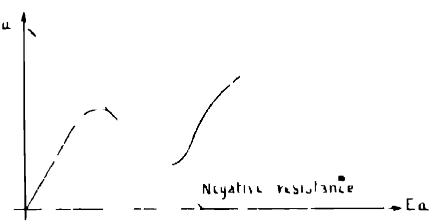


FIG 466 Anode characteristic of a tetrode showing negative resistance

(*see* Fig 466) This means that an increase in anode voltage in this region produces a decrease in anode current , in other words the AC resistance R_a is negative in this region This region occurs when the anode potential is slightly less than the screen potential

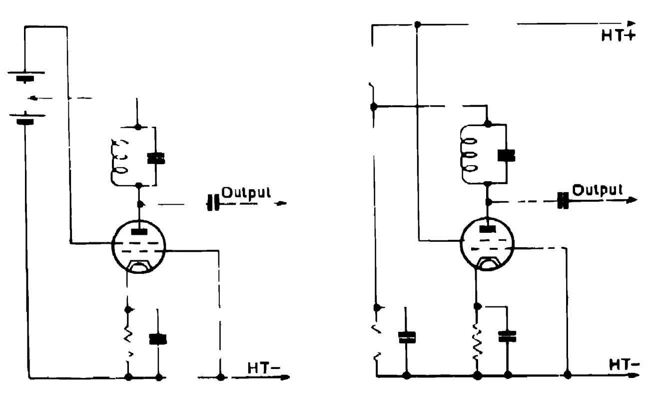


FIG 467 Dynatron oscillator
(*a*) with tap on battery
(*b*) with potential divider

Fig. 467 shows a simple dynatron oscillator. The main advantage of this type of oscillator is that, provided the negative resistance is of such a value that oscillations only just take place, the frequency stability is very high. The value of negative resistance can be adjusted by varying the bias on the grid. The reason for the good frequency stability is that the only valve characteristic likely to affect the frequency of oscillation is the anode-cathode capacity. This capacity is initially very small and the effect of any change in it due to electrode expansion will be negligible. There is in addition no variable coupling factor between the anode and grid circuits, as exists in the oscillators so far discussed.

The transitron oscillator

A pentode valve may be used as a negative resistance device. If the suppressor grid of a pentode is driven positive there is little change in the electrostatic field near the cathode and so no appreciable change in the total space current. The anode current however increases and the screen current drops. If the suppressor grid is connected to the screen *via* a condenser C' as shown in Fig. 468 there will be a negative AC resistance between the screen and earth.

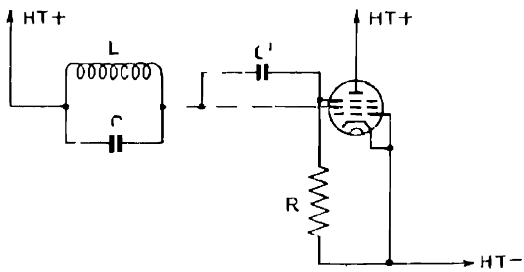


Fig. 468.—Illustrating the principle of the transitron oscillator

This may be seen from the fact that if an alternating voltage is applied to the screen on the positive half cycle the screen potential will rise, so also will that of the suppressor since they are coupled by condenser C', and this will cause the screen current to drop since the suppressor potential has a greater effect than the screen potential on the screen current. A rise in screen potential thus corresponds to a drop in screen current. On the negative half-cycle, the potential of the screen will drop, and the screen current will rise.

A parallel tuned circuit in the screen lead will therefore oscillate, being in series with a negative resistance. This is known as a transitron oscillator, and it has the same high degree of frequency stability as the dynatron type. Fig. 469 shows a typical transitron oscillator circuit. The negative resistance is controlled by means of the resistance R_1 in the suppressor circuit. A typical value for R_1 would be of the order of 25 kΩ, and this resistor is adjusted until the circuit just oscillates.

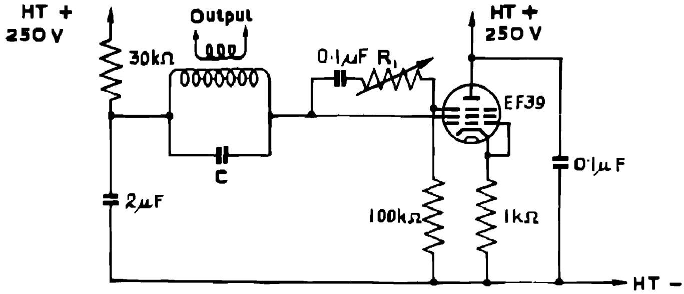


FIG. 469.—Typical transitron oscillator circuit.

Tuned-anode tuned-grid oscillator

It can be shown that, due to what is known as the "Miller effect", the anode load affects the grid-cathode impedance of a valve; moreover, the input impedance of a valve may have a negative resistance component if the anode load is inductive (*see* p. 350). This fact can be utilised to give a negative-resistance oscillator, as shown in Fig. 470.

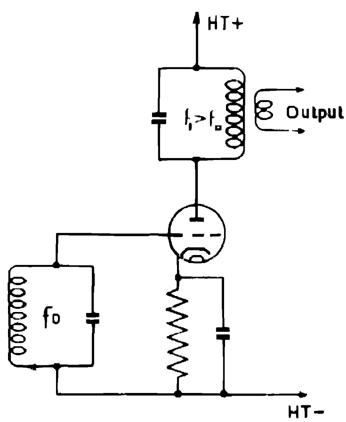
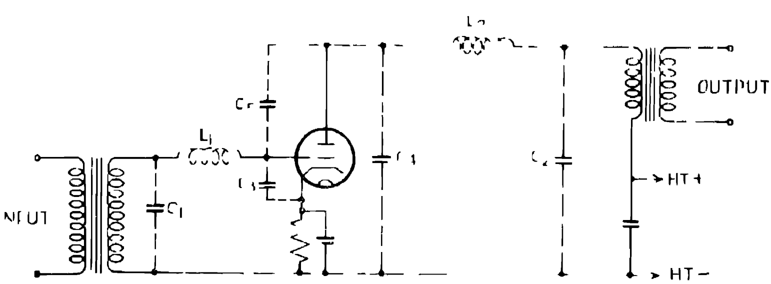


FIG. 470 Tuned-anode tuned-grid oscillator.

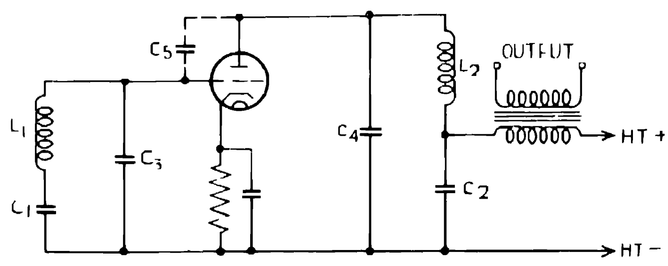
In the tuned-anode tuned-grid oscillator, the tuned circuit proper is in the grid circuit, and is tuned to a frequency f_0. The parallel circuit in the anode circuit is tuned to a frequency f_1 slightly greater than f_0, and consequently at the frequency f_0 it represents a large inductive anode load. Thus, by the effect discussed above, the tuned grid circuit will have a negative resistance across it, and oscillations at the frequency f_0 will take place. Note that there is no direct coupling between the tuned circuits. The oscillator is little used at audio-frequencies, and its most frequent application is as a high-frequency crystal-controlled oscillator, when the grid tuned circuit is replaced by a quartz crystal, the result being a very stable oscillator whose frequency corresponds to the parallel resonant frequency of the crystal.

Parasitic oscillations

High frequency oscillations sometimes occur in amplifier circuits due to stray capacity and inductance; these are known as parasitic oscillations. They are usually the result of a tuned-anode tuned-grid circuit being set up particularly if transformers are used for input and output. The tuning capacities are the valve inter-electrode capacities and these resonate with the stray inductance of the wiring. A simple amplifier circuit using transformer input and output is shown in Fig. 471*a* and Fig. 471*b* shows the equivalent circuit at high frequencies.



(*a*) Simple amplifier circuit showing stray capacities and inductances



(*b*) Equivalent circuit at high frequencies. This is a tuned anode tuned-grid oscillator

Fig. 471. Parasitic oscillations in an amplifier

The capacities C_1 and C_2 are the stray capacities to earth in the transformers and associated wiring, L_1 and L_2 are stray wiring inductances and C_3, C_4 and C_5 are valve inter-electrode capacities. C_1 and C_2 will generally be large compared with C_3 and C_4. Both grid circuit and anode circuit will have a parallel resonant frequency which will be high in view of the small capacities and inductances involved. If the parallel resonant frequency of the anode circuit is slightly higher than that of the grid circuit, the circuit will behave as a tuned anode tuned-grid oscillator, producing high frequency parasitic oscillations.

There are two main methods of combating this effect. The first is the insertion in the grid lead of a resistance sufficiently high to neutralise the negative-resistance component of the valve's input

impedance. Alternatively, a small air-core inductance may be put in series with the anode; this has the effect of increasing the stray anode inductance, and thus lowering the parallel resonant frequency of the anode circuit. If this frequency can be brought down below the parallel resonant frequency of the grid circuit, the system will not oscillate. Whether the grid resistance or the anode inductance is used, it is important to make sure that it is connected as close to the electrode as possible.

VARIABLE FREQUENCY OSCILLATORS

Coarse adjustment of frequency can be made in any tuned circuit oscillator by varying either L or C in steps. Fine adjustment is normally carried out by small continuously variable trimmer condensers across the main tuning condenser, though continuously variable inductances are sometimes used. Note that in the two cases treated mathematically on pages 471 and 479, the maintenance condition involves both L and C, and a change in frequency produced by varying L and C will also change the maintenance condition and consequently the output voltage.

Frequency coverage

An "L-C" or tuned circuit oscillator with an adjustable frequency can be made using a variable condenser for the tuning capacity C. This is quite satisfactory at radio frequencies, for although changing C will change the maintenance condition, it is possible to arrange a fairly constant output voltage over a certain working frequency range. Using normal components, this frequency range gives a ratio of approximately three to one; thus an oscillator could be made to give a frequency continuously variable from 1 Mc/s to 3 Mc/s or from 100 kc/s to 300 kc/s. Such ranges as these represent a wide variation in frequency, and such an oscillator could be useful as a test oscillator at these frequencies.

Coming down the frequency scale to the audio range, however, an audio oscillator could only be made continuously variable from, say, 300 c/s to 900 c/s or from 1000 c/s to 3000 c/s, so that a single continuously variable frequency oscillator cannot cover the whole audio range and would be of little practical use. One of the easiest solutions to this problem is the beat-frequency oscillator (BFO), which can easily be made to cover the whole audio frequency range.

The beat-frequency oscillator

Fig. 472 shows, in a block schematic form, the essential components of a beat-frequency oscillator. It consists of two high frequency oscillators, one fixed and one variable, the difference of the two frequencies being in the audio range. The two outputs are mixed together in a square-law detector or balanced modulator, and the products passed *via* one or more stages of amplification to a low-pass filter that passes only the audio component. By this

means the complete audio range can be obtained from one fixed oscillator, and one oscillator with quite a small percentage variation. For example, suppose the range required is 0 to 20 kc/s and the fixed oscillator has a frequency of 100 kc/s, then the other oscillator need only be variable between 100 kc/s and 120 kc/s. The main advantages of this type of oscillator are its simplicity and the fact that a well-designed BFO has a fairly constant output voltage over

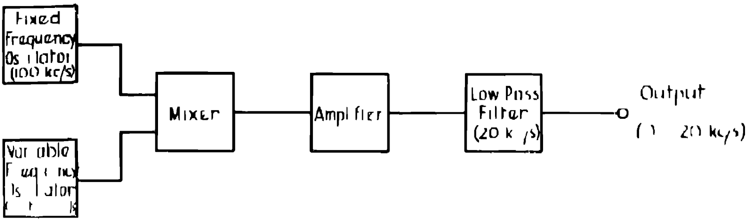


Fig. 472. Essential components of a beat-frequency oscillator.

the whole frequency range. One disadvantage is the fact that the production of harmonics is inevitable in the mixer and the output of a BFO is consequently never a perfect sine wave. Harmonic distortion is also produced in a different way. Suppose that the two high-frequency oscillators are tuned to 100 000 c/s and 100 200 c/s respectively, the fundamental in the output being 200 c/s. Since any oscillator produces harmonics, frequencies of 200 000 and 200 400 c/s are applied to the mixer producing a certain amount of 400 c/s output. This is additional to the second harmonic distortion produced by the mixer. Thus the harmonic distortion can be reduced by inserting a low pass filter with a cut-off frequency of, say, 150 kc/s, between the fixed-frequency oscillator and the mixer. This ensures that the fixed oscillator applies pure 100 kc/s to the mixer and even though the variable oscillator may be rich in harmonics there will be no harmonics in the audio output other than those produced in the mixer.

A second disadvantage is that the oscillators must be extremely stable, for a variation of 1 per cent in either of the oscillators in Fig. 472 would produce an error of 1000 c/s in the audio output. This error is largely minimised by making the two oscillators as similar as possible, so that an external influence that causes the frequency of one oscillator to vary may be assumed to cause a similar variation in the frequency of the other.

Thirdly, if a very low frequency output is attempted there is a tendency for the two oscillators to "lock" at exactly the same frequency. By using buffer amplifiers between the oscillators and the mixer, or alternatively by interposing a suitable bridge circuit, this can be largely prevented, and frequencies down to 10 c/s may be produced. Notwithstanding these facts, the BFO is excellent as a general purpose audio test oscillator.

RESISTANCE-CAPACITY OSCILLATORS

Fig. 473 shows an amplifier with a flat gain-frequency response, and zero phase-shift over a very wide frequency range. The output of this amplifier is coupled back to the input through a network whose phase-shift varies with frequency. If it can be arranged that the phase-shift of the network is zero, then the circuit will oscillate at that frequency, provided that the gain of the amplifier is equal to the loss in the network. For if this is the case, the feedback voltage is equal to the input voltage in magnitude and phase at this frequency, and this is the condition for the maintenance of oscillations.

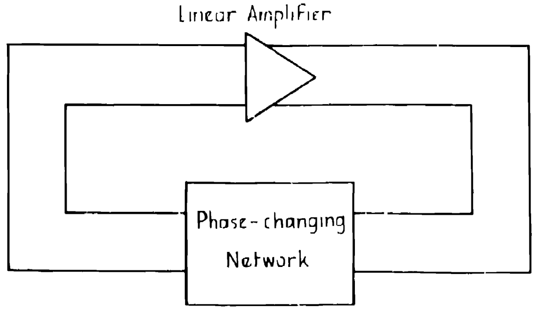


FIG. 473.—Illustrating the principle of the resistance-capacity oscillator.

Almost any combination of reactances and resistances can be used for the network. It would be an advantage, however, if the circuit could be arranged so that the required amplifier gain is independent of frequency and a network used that has a minimum attenuation at the zero phase-shift frequency. Consider the resistance-capacity network of Fig. 474.

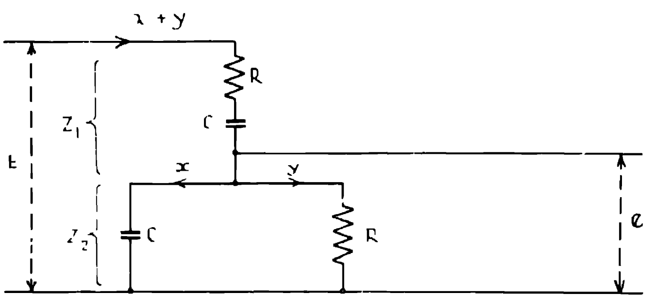


FIG. 474.—Phase-changing network suitable for use in R-C oscillators.

Let Z_1 be the impedance of R and C in series.

Let Z_2 be the impedance of R and C in parallel.

Then
$$\frac{e}{E}=\frac{Z_2}{Z_1+Z_2}$$

But $$Z_1 = R - \frac{j}{\omega C} \quad \text{and} \quad Z_2 = \frac{-\dfrac{jR}{\omega C}}{R - \dfrac{j}{\omega C}}$$

Hence
$$\frac{e}{E} = \frac{\dfrac{-\dfrac{jR}{\omega C}}{R - \dfrac{j}{\omega C}}}{R - \dfrac{j}{\omega C} - \dfrac{\dfrac{jR}{\omega C}}{R - \dfrac{j}{\omega C}}}$$

$$= \frac{-\dfrac{jR}{\omega C}}{\left(R - \dfrac{j}{\omega C}\right)^2 - \dfrac{jR}{\omega C}}$$

$$= \frac{-\dfrac{jR}{\omega C}}{R^2 - \dfrac{1}{\omega^2 C^2} - \dfrac{3jR}{\omega C}} \tag{12}$$

Thus e and E are in phase when :—

$$R^2 - \frac{1}{\omega^2 C^2} = 0$$

i.e. $$\omega^2 = \frac{1}{R^2 C^2} \tag{13}$$

or $$f = \frac{1}{2\pi RC} \tag{14}$$

Now, from equation 12 :—

$$\left|\frac{e}{E}\right|^2 = \frac{\dfrac{R^2}{\omega^2 C^2}}{\left(R^2 - \dfrac{1}{\omega^2 C^2}\right)^2 + \dfrac{9R^2}{\omega^2 C^2}} = \frac{1}{\dfrac{\omega^2 C^2}{R^2}\left(R^2 - \dfrac{1}{\omega^2 C^2}\right)^2 + 9}$$

$$\therefore \quad \left|\frac{E}{e}\right|^2 = \frac{\omega^2 C^2}{R^2}\left(R^2 - \frac{1}{\omega^2 C^2}\right)^2 + 9$$

This is a minimum when $\omega^2 = \frac{1}{R^2 C^2}$, and its value at this frequency is 9. Thus at a single frequency given by $f = \frac{1}{2\pi RC}$ the phase-shift through the network is zero and the attenuation (voltage ratio) is a minimum and equal to 3 (*see* Fig. 475). Hence for oscillations to be maintained the amplifier must have a voltage

gain of 3. The frequency is controlled by varying the two equal resistors R simultaneously by means of a single control.

Notice that from equation 14 the frequency is inversely proportional to R; it follows therefore that if resistances R_1 and R_2 give frequencies f_1 and f_2 respectively, then R_1 and R_2 in parallel will give a frequency of $f_1 + f_2$.

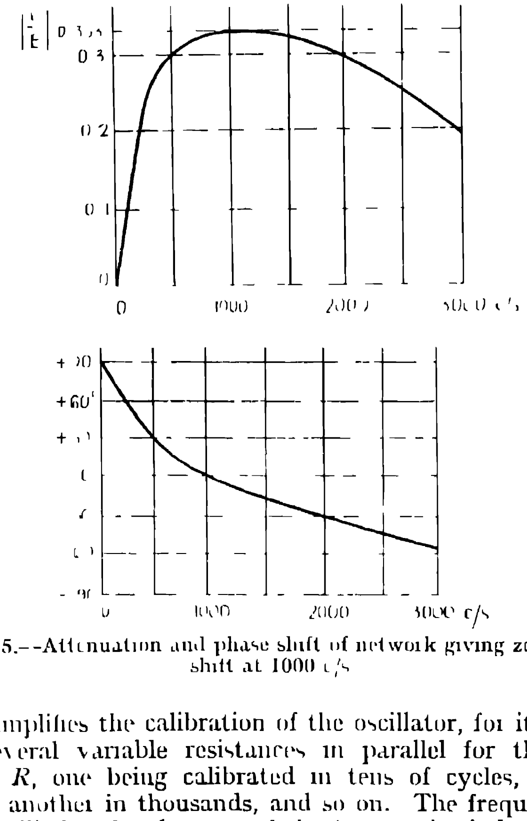


FIG. 475.—Attenuation and phase shift of network giving zero phase shift at 1000 c/s

This simplifies the calibration of the oscillator, for it is possible to use several variable resistances in parallel for the variable resistance R, one being calibrated in tens of cycles, another in hundreds, another in thousands, and so on. The frequency of the oscillator will then be the sum of the frequencies indicated on the decade controls. A further simplification is possible if C is given a value of $0{\cdot}00795$ μF. Equation 14 then reduces to $f = \dfrac{20 \times 10^6}{R}$, and this simplifies the construction of the resistance controls. It is also possible to increase the range of the oscillator by providing a "multiply by ten" key; this will merely substitute for C a capacity equal to $\dfrac{C}{10}$.

The most important part of this type of oscillator is the amplifier; for although the gain is only 3 (voltage ratio) the phase-shift must be zero at all frequencies in the working range; thus the linearity

of the amplifier is the limiting factor in the frequency range. Another point about the amplifier is that the output impedance must be low; for if this condition is not satisfied, it will be equivalent to an increase in the resistance of the series arm of the network, and consequently the calibration of the oscillator will be upset. Voltage negative feedback may be applied to obtain this low output impedance.

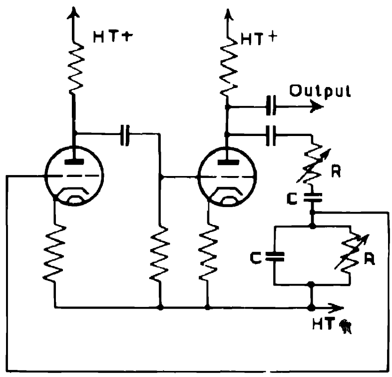


FIG. 476. Skeleton circuit of two-valve R-C oscillator

A skeleton diagram of an R-C oscillator is shown in Fig. 476, which is intended merely to show how the phase-changing network is fitted into the amplifier circuit; no details of the amplifier are shown, as this follows standard practice.

Single-stage R-C oscillator

It is possible to make a single-stage R-C oscillator, provided some network is connected between anode and grid that introduces 180° phase-shift to give positive feedback. Several networks of the type shown in Fig. 477 can be used for this purpose.

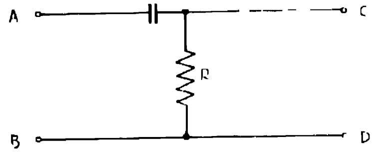


FIG. 477.—Phase-changing network

The voltage across CD in this network will be ahead of the voltage across AB by some angle between 0° and 90°, depending on the relative impedances of R and C at the frequency considered. To obtain 180° phase-shift, three of these networks have to be used, as it is not possible to get quite 90° phase-shift from one.

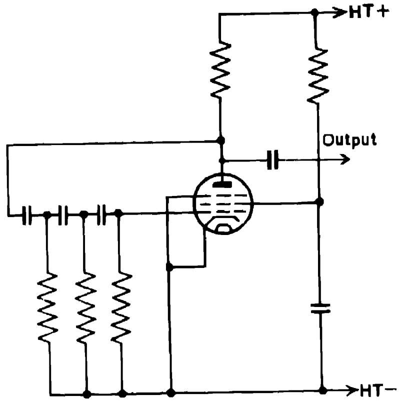


Fig. 478 Single stage R C oscillator

The three networks are connected between anode and grid, giving a circuit of the type shown in Fig. 478. This type of oscillator is most suitable for use at fixed frequencies.

THE MULTIVIBRATOR

The most common relaxation oscillator circuit, and the only one to be considered here, is the multivibrator, shown in Fig. 479.

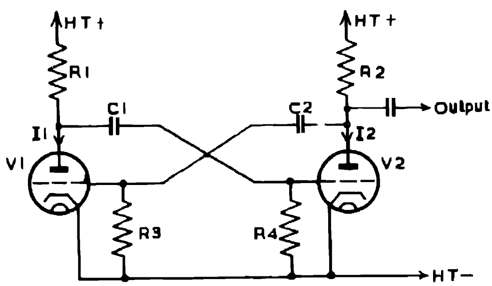


Fig. 479 Multivibrator circuit

It is assumed for simplicity that $R_1 = R_2$, $R_3 = R_4$ and $C_1 = C_2$, and this is usually the case in practice. It will be seen that the multivibrator consists of a two-stage resistance-capacity coupled amplifier, with the output fed directly back to the input. As the total phase-shift through such an amplifier is 360°, clearly the circuit will oscillate, the frequency depending on the component values.

To follow the operation of the circuit, suppose that, on switching on, a small positive voltage appears on the grid of V_1. This causes an increase in the anode current I_1 of the valve V_1, and a decrease in the anode potential of V_1 which is passed to the grid of V_2. This decrease in the grid potential of V_2 causes a decrease in I_2,

and an increase in the anode potential of V_2; this drives the grid of V_1 more positive. This process is continued and the effect is cumulative, the grid of V_1 growing more positive and the grid of V_2 more negative until V_2 is biased back to cut-off; the whole process is, of course, practically instantaneous. The circuit remains in this condition while the negative potential on the grid of V_2 leaks away, the time for this depending on the time constant of C_1 and R_4. The positive potential on the grid of V_1 disappears much more quickly due to the flow of grid current. As soon as the grid potential of V_2 reaches the point where anode current can flow once more, the anode potential of V_2 will begin to fall due to the fact that I_2 is increasing; this drives the grid of V_1 negative and the whole action is repeated in the opposite direction until V_1 is cut off. The grid potential and anode current waveforms for each valve are shown in Fig. 480.

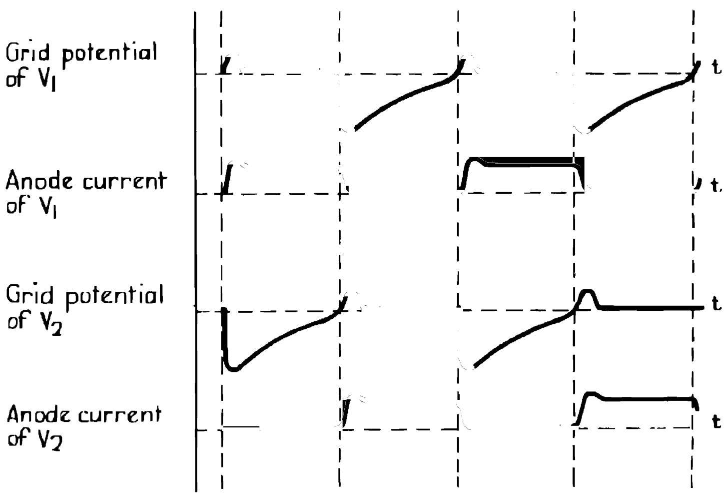


FIG. 480.—Variation of grid potential and anode current in multivibrator.

These waveforms are particularly rich in harmonics, the anode currents being approximately square wave, and this is one of the important properties of the multivibrator. The second important property is the fact that the frequency can easily be locked in synchronism with another frequency. The best method if pentodes are used is to inject the synchronising voltage on to the suppressor grid of either valve. The multivibrator will lock even if the injected frequency is several times its own frequency. Suppose, for example, a multivibrator is working at a frequency of the order of 100 cycles per second. If a 1000 c/s voltage is injected, the multivibrator can be made to lock at 100 c/s. In this way a frequency dividing system is obtained.

The frequency can be varied by altering R_3 and R_4, and may be as low as desired by making these resistances large enough (*e.g.*, one

cycle per minute is quite possible with this circuit using very high values for R_3 and R_4). At the other end of the scale, the limit on high frequencies is about 100 kc/s, which is about the limit for satisfactory resistance-capacity coupling.

It was assumed for simplicity that $R_1 = R_2$, $R_3 = R_4$, and $C_1 = C_2$; this simplification gives a symmetrical output, the valves being cut off for equal periods of time. In the more general case when these equalities do not hold, the operation of the circuit is just the same, but the valves are cut off for unequal periods, and the output is asymmetrical.

CHAPTER 11

MODULATION

ADDITION OF TWO SINE WAVES

If two pure sine waves are applied simultaneously to a communication system, their instantaneous amplitudes may be added. Fig. 481 shows the equivalent circuit of a network in which the sum of the two voltages e_1 and e_2 appears across the impedance Z_1; Fig. 482 shows the equivalent circuit of a network in which the sum of the currents i_1 and i_2 flows through the impedance Z_2.

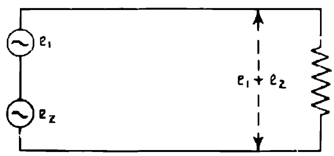


FIG. 481.—Addition of two sinusoidal voltages.

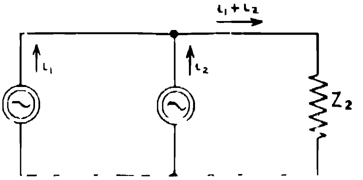


FIG. 482.—Addition of two sinusoidal currents.

Consider the two sine waves shown in Fig. 483*a* and *b*. These waveforms have equal amplitudes, and their frequencies f_1 and f_2 differ only slightly. The addition of these two waveforms is given by Fig. 483*c*. This resultant waveform *c* consists only of frequencies f_1 and f_2; filters, tuned circuits, or any similar simple device will be sufficient to re-separate the independent frequencies. Applying Fourier's analysis to the waveform verifies the fact that no other frequencies are present.

Curve *c* may, however, be considered as a waveform having a frequency $\frac{f_1 + f_2}{2}$ and varying in amplitude in such a way as to have a cosine-wave envelope of frequency $\frac{f_1 - f_2}{2}$. This may be seen from Fig. 483, and verified mathematically.

Let $\quad y_1 = A \,.\, \sin 2\pi f_1 t$

and $\quad y_2 = A \,.\, \sin 2\pi f_2 t$

Then $y_1 + y_2 = A\,(\sin 2\pi f_1 t + \sin 2\pi f_2 t)$

$$= 2A \sin 2\pi \frac{f_1 + f_2}{2} t \,.\, \cos 2\pi \frac{f_1 - f_2}{2} t$$

This is equivalent to a sine wave of frequency $\frac{f_1 + f_2}{2}$ varying cosinusoidally in amplitude from $+2A$ through 0 to $-2A$, and back through 0 to $+2A$ again, with a frequency of $\frac{f_1 - f_2}{2}$. If the *absolute* amplitude of the waveform is considered, the increase and decrease occurs at a frequency of $(f_1 - f_2)$, the latter is called the "beat frequency."

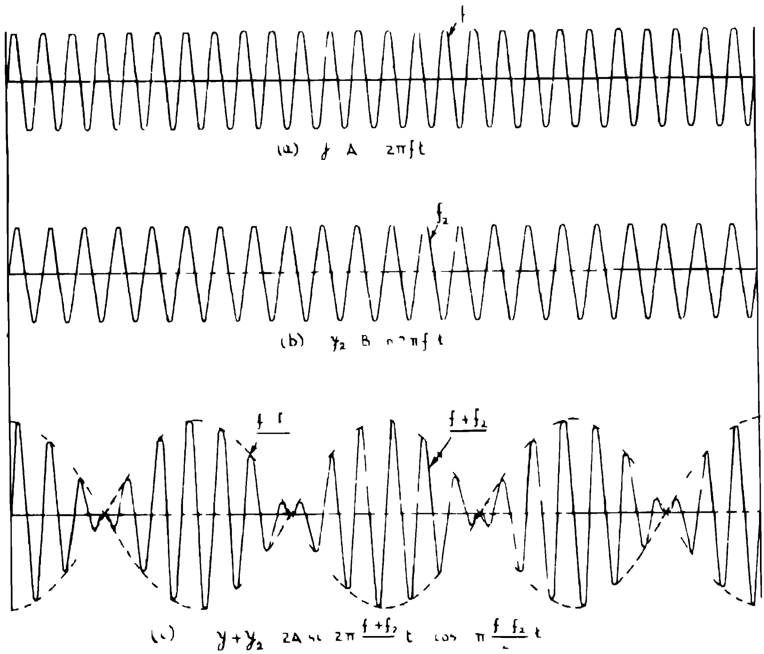


FIG. 483. Addition of two sine waves of equal amplitudes (i.e. $A = B$).

Beat notes

The above discussion applies to all sinusoidal waveforms and hence it applies equally well to sound waves. This is important, because it gives rise to *audible* beat notes. Suppose that the frequencies under consideration are $f_1 = 255$ c/s and $f_2 = 257$ c/s. These two, when present simultaneously, will produce exactly the same sound as a 256 c/s frequency varying in amplitude in a cosinusoidal manner with a frequency of 2 c/s, the two resultant waveforms are in fact exactly equivalent.

When frequencies of this nature are close together and in the audio range the human ear cannot distinguish the individual frequencies, and the tones of 255 c/s and 257 c/s, when present simultaneously, give the impression that a note of 256 c/s is present, varying in amplitude at a frequency of 2 c/s. This is called a 2 c/s

beat note. If the frequencies were 254 c/s and 258 c/s the ear would give the impression that only a 256 c/s note were present varying in amplitude at a frequency of 4 c/s, and so on.

However, this state of affairs applies only when the frequencies are close together within, say 30 c/s. Listening simultaneously to notes of 286 c/s and 226 c/s one hears the two pure sine waves, not a note of 256 c/s "beating" at 60 c/s.

Waves of different amplitudes

So far it has been assumed that the amplitude of the sine wave of frequency f_1 is the same as that of f_2. If the amplitudes are

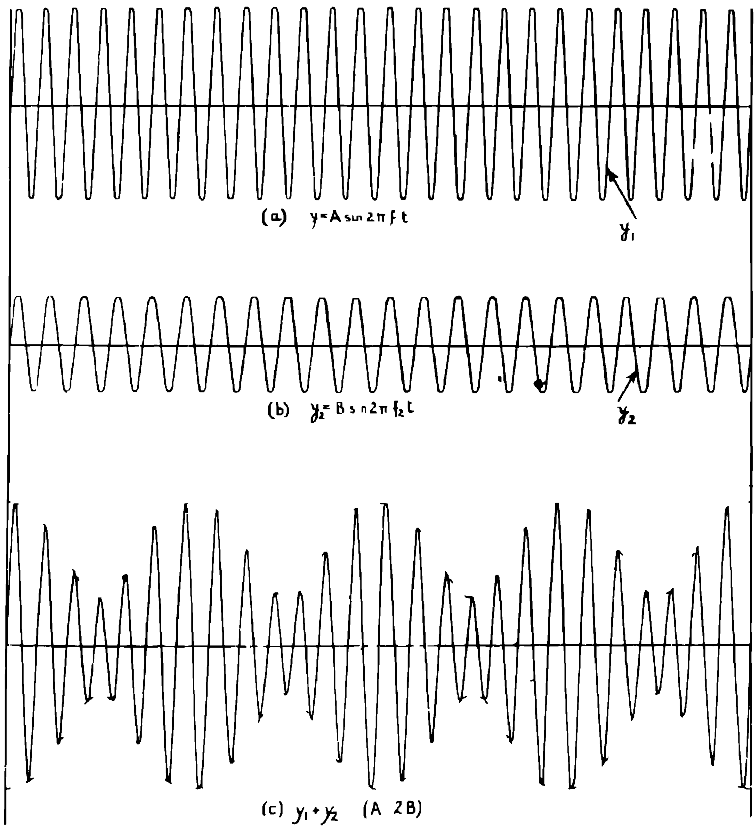


Fig. 484.—Addition of two sine waves of unequal amplitudes ($A = 2B$)

different, the general shape of the resultant will alter slightly, but the beat note may still be detected.

Fig. 484 shows the addition of two sine waves when the amplitude of y_1 is twice the amplitude of y_2.

Synchronising two audio frequencies

This " beat note " phenomenon gives a very simple method of synchronising two frequencies within the audio range, but it should be remembered that no beat note will be heard if the parent frequencies are outside the audio limits. In such a case an AC meter would give a convenient method of detecting the variation in amplitude due to the presence of beat notes, since the deflection obtained is a function of the envelope amplitude of the waveform.

Waves of widely different frequencies

This general shape of waveform is maintained until the frequency of y_1 becomes exactly twice that of y_2. When this occurs,

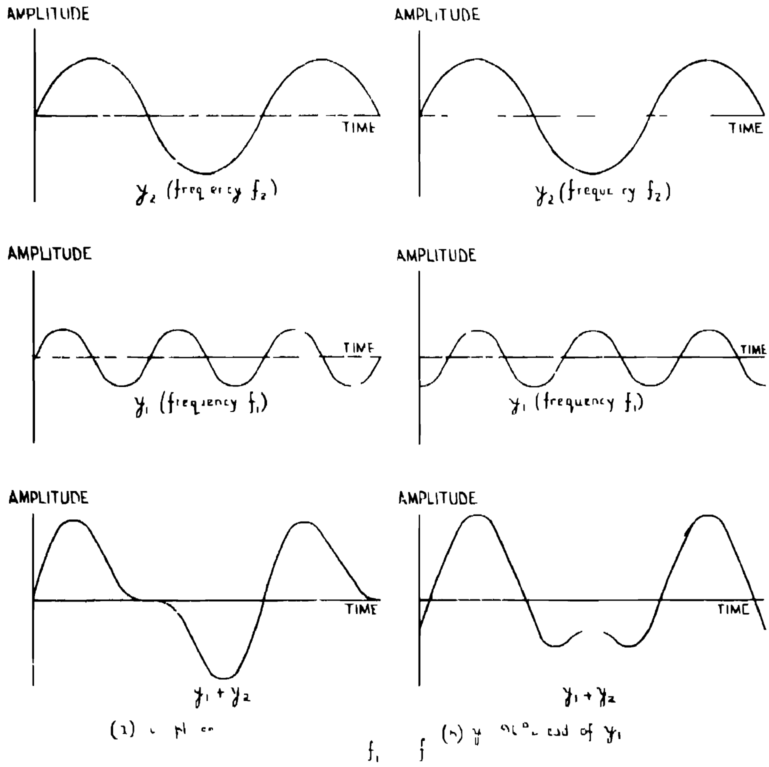
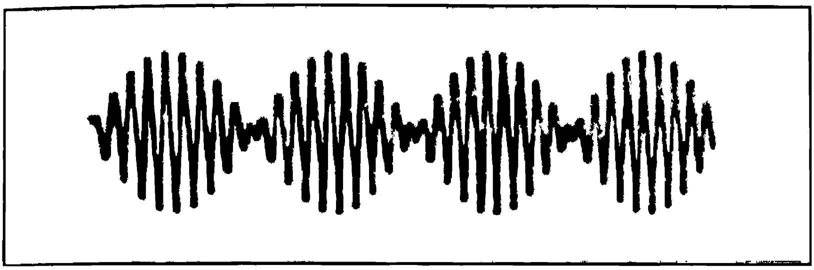


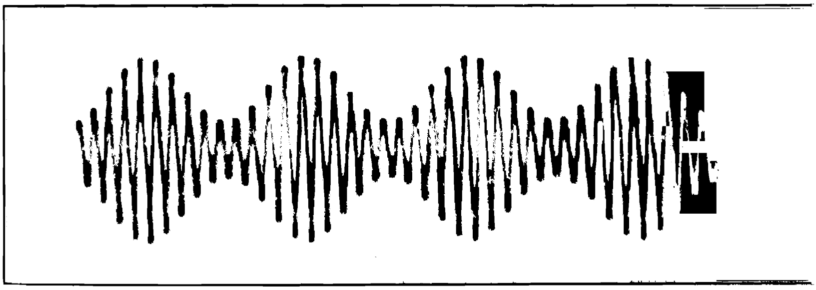
Fig. 485 — Addition of two sine waves of different amplitudes, the frequency of one being twice that of the other

the waveform is simply that of a sine wave (of frequency f_2) with second harmonic distortion, the exact shape of the resultant depending on the relative phases of y_1 and y_2 (*see* Fig. 485).

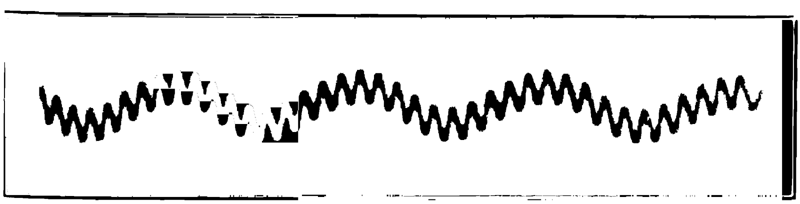
When the frequency of y_1 is greater than twice that of y_2, the resultant waveform shows clearly the superposition of y_1 on y_2.



(*a*) Addition of two sine waves of equal amplitude



(*b*) Addition of two sine waves of unequal amplitude.



(*c*) Addition of two sine waves the frequency of one being ten times the frequency of the other

PLATE 27—CRO traces

Fig. 486*a* shows the addition waveform of the sine waves y_1 and y_2 with the amplitude of y_1 equal to one-tenth of the amplitude of y_2. Fig. 486*b* shows the resultant of the two curves when the amplitudes are equal. In both cases the frequency of y_1 is twenty times the frequency of y_2. The two resultants should be contrasted with those of Figs. 483 and 484.

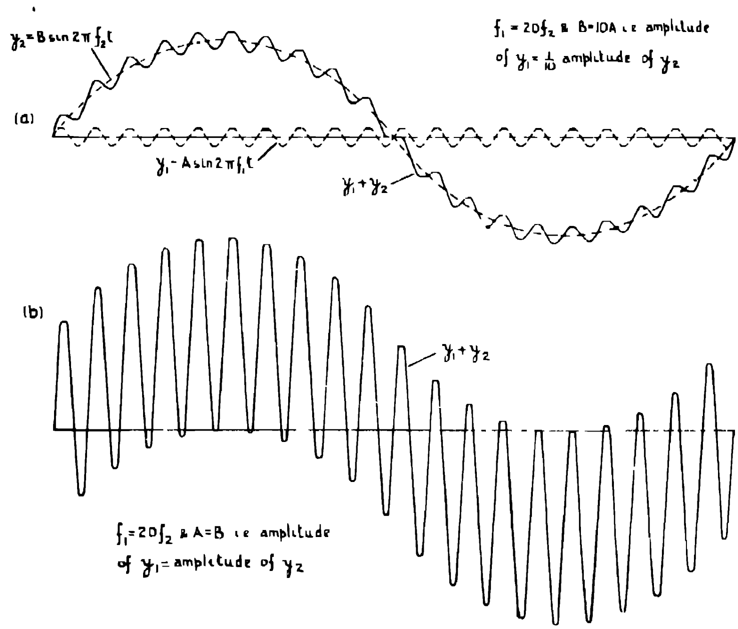


FIG. 486.- Addition of two sine waves, the frequency of one being twenty times the frequency of the other.

AMPLITUDE MODULATION

If the amplitude of a sine wave of frequency f_c is varied sinusoidally between the limits $A + a$ and $A - a$ with a frequency f - *i.e.*, the amplitude at any instant is $(A + a \,.\, \sin 2\pi ft)$ —then the equation of the resultant waveform will be : -

$$y = (A + a \,.\, \sin 2\pi ft) \,.\, \sin 2\pi f_c t$$

i.e.
$$y = A\left(1 + \frac{a}{A} \sin 2\pi ft\right) . \sin 2\pi f_c t \qquad (1)$$

Such a waveform is known as an "amplitude modulated" waveform ; the frequency f_c is known as the "carrier" frequency, and the frequency f is known as the "modulating" frequency ;

$\frac{a}{A}$ is known as the "modulation factor" and $\frac{100a}{A}$ per cent. as the "percentage modulation" (*see* Fig. 487).

Equation 1 for the waveform may be simplified by letting

$$2\pi f = \omega$$
$$\text{and} \quad 2\pi f_c = p$$

In this case equation (1) becomes:—

$$v = A\left(1 + \frac{a}{A}\sin \omega t\right)\sin pt \qquad (2)$$

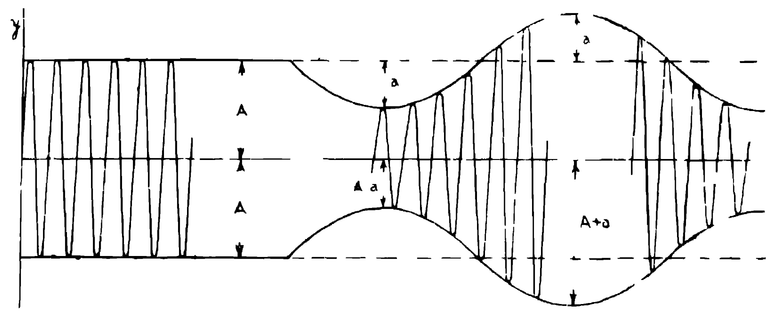


Fig. 487. Amplitude modulation.

Sidebands

This resultant waveform does not consist simply of frequencies f_c and f as would be the case had the two frequencies been "added", in fact there is no longer any component of frequency f present in the resultant although two new frequencies $f_c + f$ and $f_c - f$ have been produced.

This may be verified mathematically:

$$A\left(1 + \frac{a}{A}\sin \omega t\right)\sin pt = A\sin pt + a\sin pt \sin \omega t$$

$$= A\sin pt + \frac{a}{2}\cos(p - \omega)t - \frac{a}{2}\cos(p + \omega)t \qquad (3)$$

(from equation 40, p. 43)

It will be noted that the amplitude a and the frequency f $\left(=\frac{\omega}{2\pi}\right)$ of the modulating waveform appear only in the second and third terms.

Considering the terms separately:

(*a*) $A \sin pt$ corresponds to a sine wave of the carrier frequency (f_c) and of maximum amplitude A.

(*b*) $\frac{a}{2}\cos(p - \omega)t$ corresponds to a sine (or cosine) wave of frequency $(f_c - f)$ and of maximum amplitude $\frac{a}{2}$. This frequency is known as the "lower sideband" (LSB) frequency.

(c) $-\frac{a}{2}\cos(p+\omega)t$ corresponds to a sine (or cosine) wave of frequency $(f_c + f)$ and of maximum amplitude $\frac{a}{2}$. This frequency is known as the "upper sideband" (USB) frequency.

It follows, therefore, that an amplitude modulated waveform is equivalent to three sine waves, namely the carrier, the upper sideband, and the lower sideband.

Thus the waveform, shown in Fig. 487, may be separated by filters or tuned circuits into its three constituent sine waves having frequencies f_c, $(f_c + f)$, $(f_c - f)$.

Conversely, three sine waves having the correct amplitudes, frequency and phase will, when present simultaneously, be equivalent to an amplitude modulated waveform.

Example.—

If a carrier frequency of 6 kc/s is amplitude modulated by an audio frequency of 1600 c/s, what frequencies will be present in the output ?

Carrier	6 kc/s
Upper sideband	$= 6 + 1{\cdot}6 = 7{\cdot}6$ kc/s
Lower sideband	$= 6 - 1{\cdot}6 = 4{\cdot}4$ kc/s

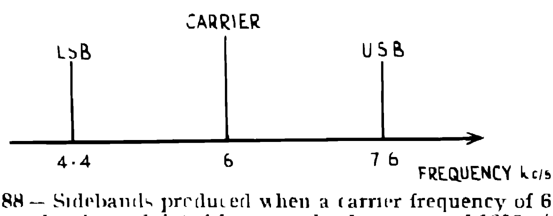


FIG. 488.—Sidebands produced when a carrier frequency of 6 kc/s is amplitude modulated by an audio frequency of 1600 c/s.

If the carrier frequency is modulated simultaneously by a number of frequencies a number of sideband frequencies will be produced.

Example.—

If a carrier frequency of 300 c/s is modulated by a square-waveform of 25 c/s, what frequencies will be produced ?

A square waveform of 25 c/s may be considered as the sum of an infinite number of sine waves (p. 106), all of which are odd harmonics of a 25 c/s sine wave, the amplitude decreasing as the number of the harmonic increases. The modulating frequencies present are therefore 25 c/s, 75 c/s, 125 c/s, 175 c/s, *etc.*, with decreasing amplitudes, up to infinity.

These will give rise to a series of upper and lower sidebands,

thus :—

Lower Sidebands.	*Carrier.*	*Upper Sidebands.*
	300 c/s	
275 c/s		325 c/s
225 c/s		375 c/s
175 c/s		425 c/s
125 c/s		475 c/s
. . .		. . .

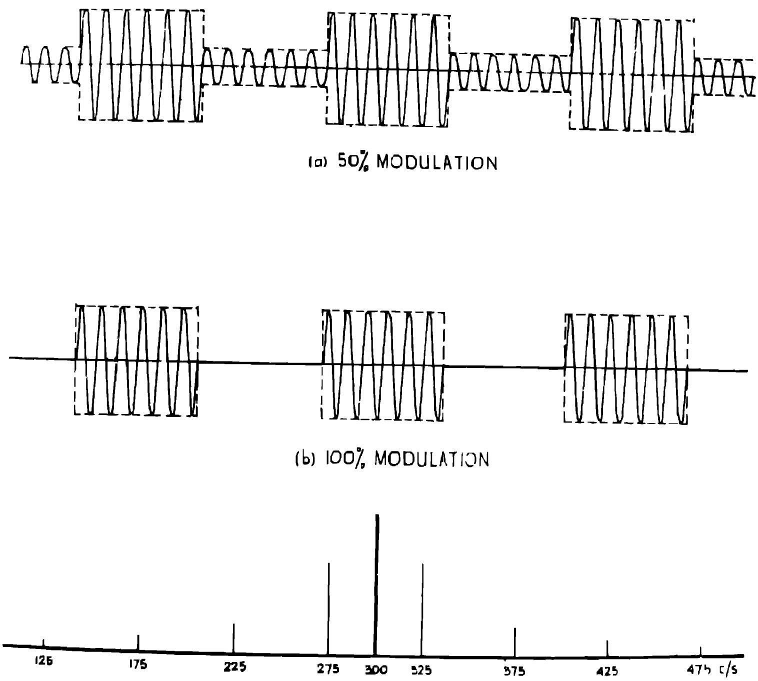


FIG. 489.—Modulation of 300 c/s with a square-waveform, frequency 25 c/s.

It should be noted that 100 per cent. modulation with such a square-waveform is equivalent to switching the carrier on and off at a rate of 25 times a second.

If the carrier frequency is modulated by a band of frequencies, only the highest and lowest frequencies in the band are usually considered, and the sideband frequencies produced by these are calculated. All other sideband frequencies will be within these limits.

Example.—

A carrier frequency of 16 kc/s is modulated by audio frequencies ranging from 300 to 2700 c/s. What will be the range of the upper and lower sidebands?

300 c/s will produce an upper sideband frequency of 16·3 kc and a lower sideband frequency of 15·7 kc/s.

2700 c/s will produce an upper sideband frequency of 18·7 kc/s and a lower sideband frequency of 13·3 kc/s.

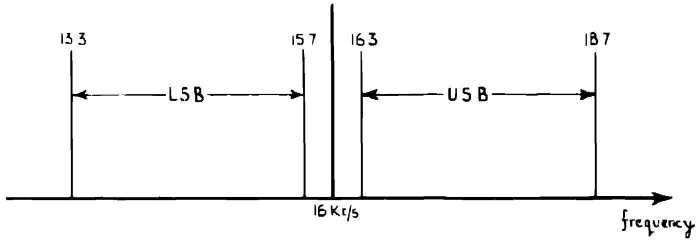


FIG. 490 Sidebands produced when a carrier frequency of 16 kc/s is amplitude modulated by audio frequencies of 300 c/s and 2700 c/s.

Thus the upper sideband will range from 16·3 to 18·7 kc/s, and the lower sideband will range from 15·7 down to 13·3 kc/s, as shown in Fig. 490.

Application of modulated waveforms

The chief function of a modulated waveform is to transmit the intelligibility originally contained in the modulating signal, but in a different frequency band. The transmission of intelligibility entails the transmission of signals corresponding to both the amplitude and the frequency of the modulating waveform. It may be conveyed in the following ways:—

(1) By transmitting the complete amplitude-modulated signal, consisting of carrier, upper sideband, and lower sideband,

(2) by transmitting the carrier and the upper sideband only,

(3) by transmitting the carrier and the lower sideband only,

(4) by transmitting the upper sideband alone, and replacing the carrier at the receiving station,

(5) by transmitting the lower sideband alone and replacing the carrier at the receiving station.

In all cases the original modulating signal is regained at the receiving station by a process known as "demodulation". All the above methods are employed in line communication systems, and will be discussed more fully in Chapter 22.

Since the intelligibility is contained in the sidebands, it is useless to transmit the carrier alone. The only remaining method would be to suppress the carrier and to transmit both the upper and lower sidebands together. This method has not been included in the above list, since it is seldom used owing to the practical difficulties of demodulating the signal at the receiving end. In such a case,

the carrier must be replaced at the receiving station with not only exactly correct frequency, but also the correct phase.

MODULATORS

The function of a modulator is to produce sidebands. As has been seen, the exact composition of the signal transmitted, and hence the modulator used, will depend on the method of working adopted. The modulators now to be described all deliver, in their output, the carrier frequency as well as both upper and lower sidebands, and they illustrate the fundamental principles of amplitude modulation. After these, will be described the modulators used when it is desired not to transmit the carrier.

Non-linear impedance modulator

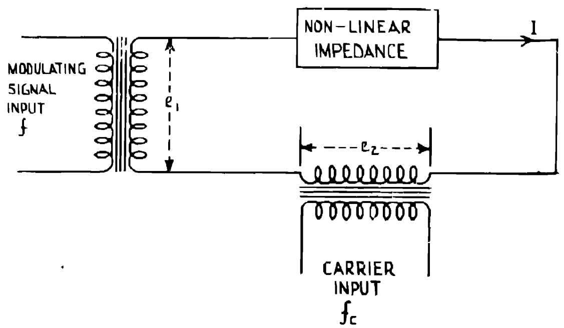


FIG. 491.—Simple non-linear impedance modulator.

Fig. 491 shows the circuit of a simple non-linear impedance (NLI) modulator. Assume that the non-linear impedance has a characteristic such that :—

$$i = a + b \,.\, e + c \,.\, e^2$$

where e is the voltage applied, i is the current flowing, and a, b and c are constants.

Let the current flowing be I,

and the voltage applied $= e_1 + e_2$

$$= E_1 \sin \omega t + E_2 \sin pt$$

$$\text{Hence } I = a + b\,(E_1 \sin \omega t + E_2 \sin pt) + c\,(E_1 \sin \omega t + E_2 \sin pt)^2$$

$$= a + bE_1 \sin \omega t + bE_2 \sin pt + cE_1E_2 \cos (p - \omega)t$$
$$- cE_1E_2 \cos (p + \omega)t + \text{harmonics.}$$

= DC + modulating signal + carrier + LSB + USB + harmonics.

If it is desired to obtain an output voltage containing the modulation products, a linear impedance Z must be included in the circuit, and the voltage taken off from across Z.

Thermionic valves and metal rectifiers as modulators

Although *any* non-linear impedance may be employed as modulator, there are two main classes of modulators in general use:

(1) *Thermionic valve modulators,* normally employed when the complete modulated waveform is to be transmitted.

(2) *Metal rectifier modulators,* normally used when a "suppressed carrier" output waveform is required.

Valve modulators

The circuits employing valve modulators may be divided into two classes. In the first, the modulating frequency and the carrier frequencies are both applied between the same grid and the cathode (see

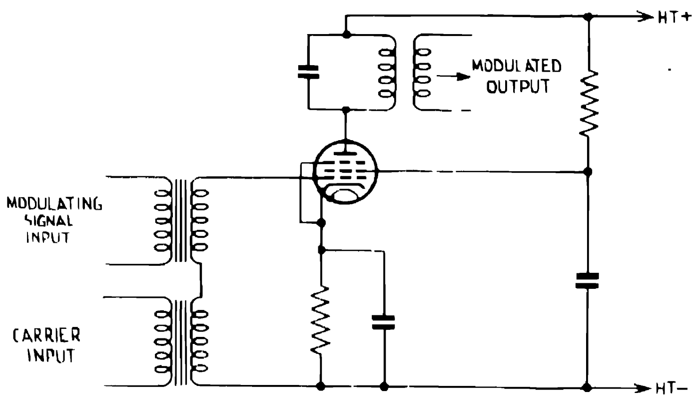


Fig. 492.—Thermionic valve modulator using grid modulation.

Fig. 492), and the non-linear part of the mutual characteristic of the valve is utilised to produce the required modulation products. In this respect the valve behaves in a similar manner to the non-linear impedance just discussed, except that the amplification property of the valve is utilised in addition to its non-linear property. This method is known as "grid modulation".

The disadvantage of this method of modulation is the interaction that tends to occur between the carrier and signal input circuits.

In the second class, separate electrodes are used for the injection of the two frequencies, and reliance is placed upon the fact that the voltage applied to one grid produces a variation in the slope of the mutual characteristic of the other. Since the only coupling between the two signals is *via* the electron stream, interaction is reduced to a minimum. However, this method entails the use of a multi-grid valve, of which the pentode is a simple type. Fig. 493 shows the use of a pentode valve as a modulator, the modulating signal being applied to the control grid, and the carrier frequency applied to the suppressor. This method is known as suppressor grid modulation; the circuit is taken from a high frequency carrier telephone

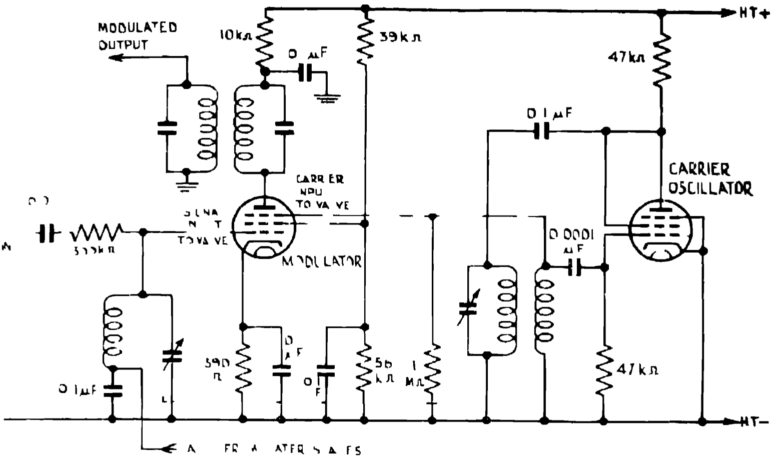


Fig 493 —Suppressor grid modulation

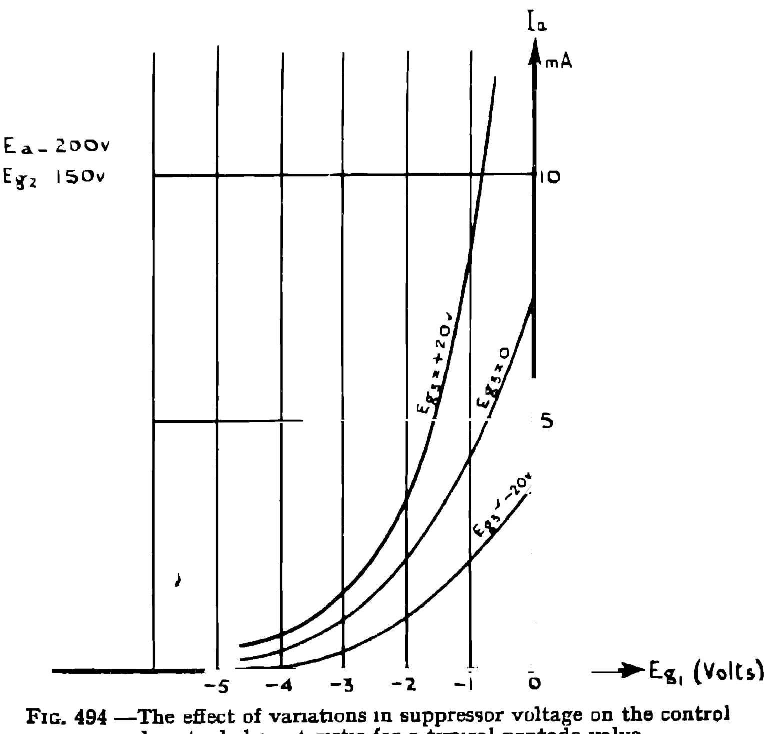


Fig. 494 —The effect of variations in suppressor voltage on the control grid mutual characteristic for a typical pentode valve.

system. Fig. 494 shows the effect of variations in suppressor voltage on the control grid mutual characteristic for a typical pentode valve.

As has been seen in Chapter 7, a number of different types of valve have been designed specially for use as modulators. In

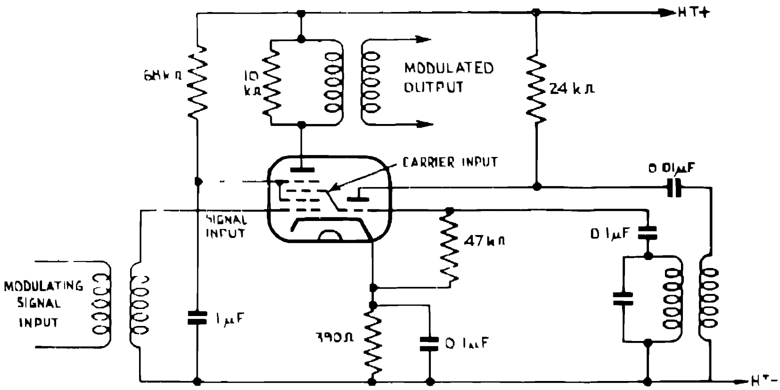


FIG. 495.—Modulation using a triode-hexode frequency changer.

certain cases the triode oscillator producing the carrier frequency may be included in the same envelope. Fig. 495 shows a modulator circuit employing a valve of this type—a triode-hexode.

Metal rectifier modulators

In carrier telephone circuits, where the carrier frequency is low (below 40 kc/s in a three-channel carrier telephone system), metal rectifier modulators may be used successfully. At higher frequencies the inherent capacity of the rectifiers introduces a shunt loss, and unless care is taken to reduce this loss to a minimum, satisfactory working will be impossible. Metal rectifier modulators employing low-capacity rectifier elements have nevertheless been devised for operation at carrier frequencies up to 2 Mc/s. Above this frequency, thermionic valve modulators are used.

Metal rectifier modulators are low-level devices, and will operate only at a part of a circuit where the power level is low. This means that, in general, such a metal rectifier modulator will be followed by a power amplifier.

MODULATION IN SUPPRESSED CARRIER SYSTEMS

The output of all the modulators so far described consists of both upper and lower sidebands and the carrier. The "balanced" modulators now to be considered "suppress" the carrier, so that

their output contains no component of carrier frequency, but only the two wanted sidebands, plus some unwanted products of modulation, which can be removed by filters.

Balanced modulators

A balanced modulator is one that gives an output containing the upper and lower sidebands, but with the carrier suppressed. The presence of these two sidebands simultaneously will give rise to an output corresponding to a beat note waveform.

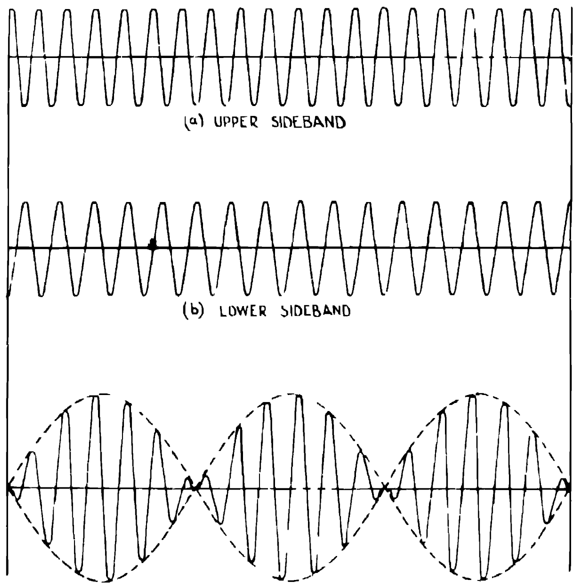


FIG 496 —Addition of two sidebands without carrier.

Fig. 497*a* shows a typical balanced modulator circuit. The carrier is applied to the centre-taps of transformers $T1$ and $T2$, and provided that rectifiers $W1$ and $W2$ are exactly matched, no carrier frequency component will be present in the output of $T2$. A "carrier leak" control may be provided by the insertion of a potentiometer at B, as shown in Fig. 497*b* This potentiometer enables the relative magnitudes of the carrier currents flowing through the two halves of the primary of $T2$ to be adjusted until the minimum carrier output is obtained.

The operation of this balanced modulator will now be explained. It has been seen that the impedance of a metal rectifier to small AC signals varies considerably with the DC bias voltage applied (*see* p. 293), being very small when the rectifier is forward-biased, and very large when the rectifier is back-biased (*see* Fig. 497*c*).

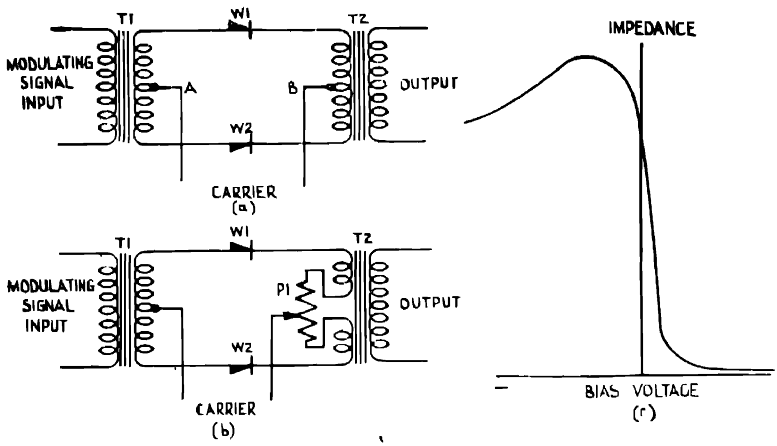


FIG. 497.—Balanced modulator.

Assuming that the amplitude of the carrier input is large compared with the modulating signal input, only the biasing of the rectifiers by the carrier need be considered. Consider the half-cycles of the carrier when A is at a higher potential than B (Fig. 497*a*): the rectifiers will be forward-biased (*see* Fig. 498), the path between the two transformers $T1$ and $T2$ will have a low impedance, and the signal tone will pass to the output. During the other half-cycle when B is at a higher potential than A, the rectifiers are back-biased (*see* Fig. 499) and the impedance between $T1$ and $T2$ will be high, preventing signal tone from passing to the output.

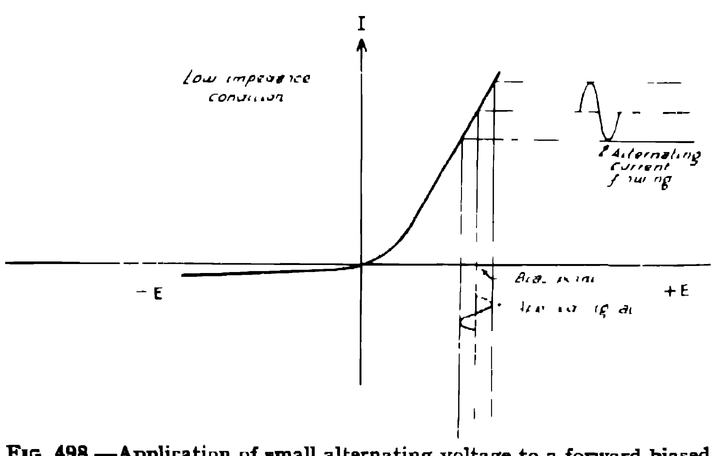


FIG. 498.—Application of small alternating voltage to a forward-biased metal rectifier.

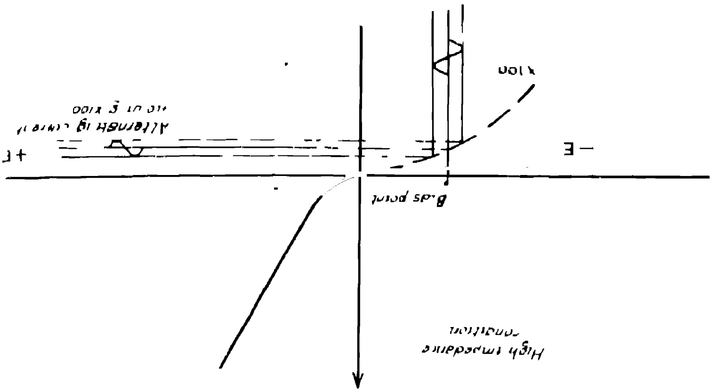


FIG. 499.—Application of small alternating voltage to a back-biased metal rectifier.

The input and output waveforms are shown in Fig. 500 The output waveform appears to show no resemblance to the anticipated waveform of Fig. 496c. This is due, however, to the presence of a large number of frequencies, including a modulating signal component in addition to the upper and lower sidebands. The output waveform follows the modulating signal input during the positive half-cycles of the carrier, but is zero during the negative half-cycles. This is exactly the waveform produced when the signal input waveform of Fig. 500a is multiplied by the square waveform of Fig. 500d.

$$\text{Output} = E_1 \sin \omega t \times \left[\tfrac{1}{2} + \frac{2}{\pi}\{\sin pt + \tfrac{1}{3}\sin 3pt + \dots\}\right]$$

$$= \frac{E_1}{2}\sin \omega t + \frac{2E_1}{\pi}\sin pt \, . \sin \omega t + \frac{2E_1}{3\pi}\sin 3pt \sin \omega t + \dots$$

$$= \frac{E_1}{2}\sin \omega t + \frac{E_1}{\pi}\left[\cos (p - \omega)t - \cos (p + \omega)t\right]$$

$$+ \frac{E_1}{3\pi}\left[\cos (3p - \omega)t - \cos (3p + \omega)t\right] + \dots$$

= modulating signal + upper and lower sidebands of carrier frequency + upper sideband and lower sideband of 3 times carrier frequency + . . .

The output thus consists of a modulating signal component and the wanted upper and lower sidebands, plus an infinite number of sidebands corresponding to the odd harmonics of the carrier, the amplitude decreasing as the number of the harmonic increases. Note that there are no sidebands present corresponding to any harmonics of the modulating signal.

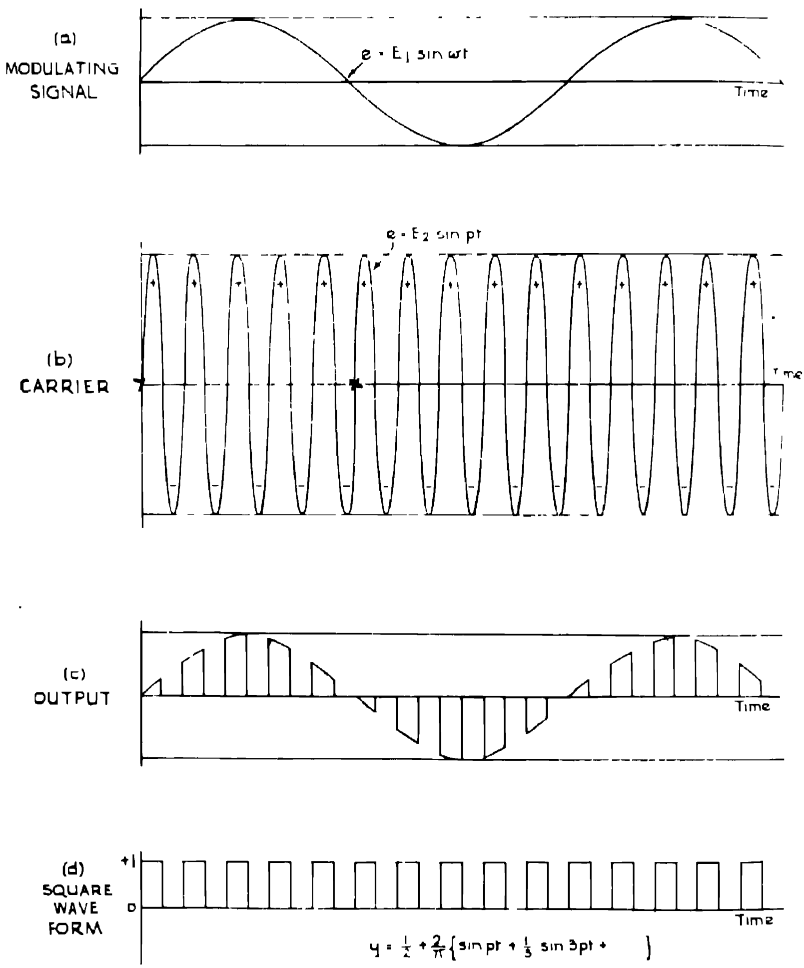


FIG. 500. —Waveform produced by balanced modulator.

The Cowan modulator

An alternative form of balanced modulator frequently used in America is the Cowan modulator shown in Fig. 501.

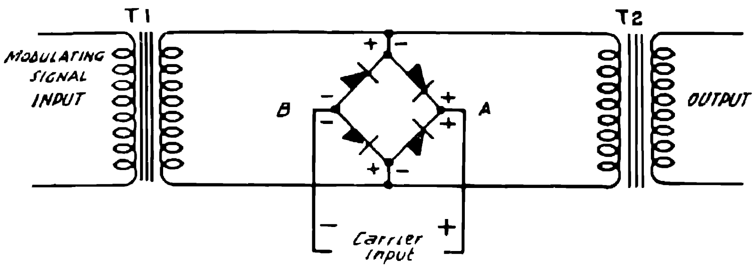


FIG. 501.—The Cowan modulator.

Its operation may be explained as follows. During those half-cycles of the carrier when A is at a higher potential than B, as indicated in Fig. 501, all the rectifiers are back-biased, and the rectifier network presents only a negligibly small shunt loss in the signal path from $T1$ to $T2$ (*see* Fig. 502*a*).

During those half-cycles of carrier when B is at a higher potential than A, all the rectifiers are forward-biased, and the rectifier network presents a virtual short-circuit across the signal path,

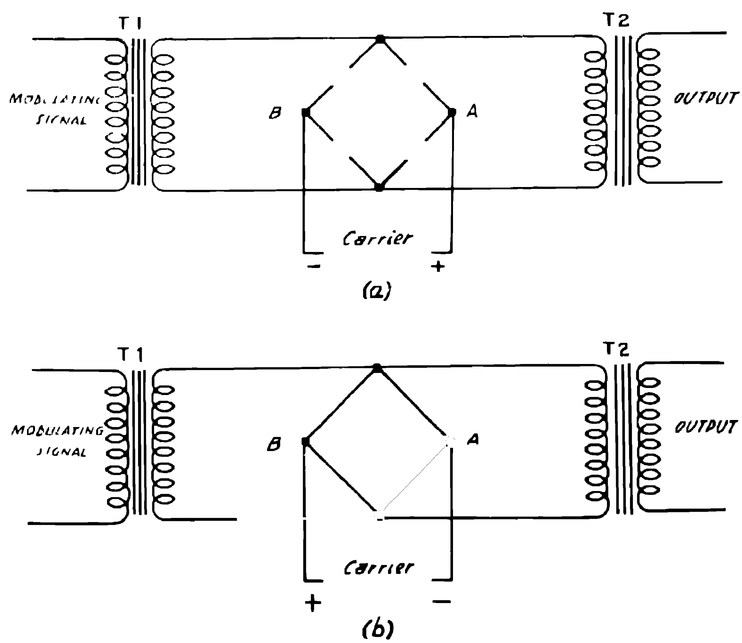


FIG. 502.—Equivalent circuits for the Cowan modulator.

preventing tone from reaching the output. This will give rise t, exactly the same output waveform as that shown in Fig. 500r, and the output will contain the modulating signal, upper and lower sidebands, and additional modulation products, but no carrier

Double-balanced bridge-ring modulators

It is possible to design modulators in which not only the carrier but also the modulating signal frequencies are suppressed. The suppression of the carrier is achieved in a manner similar to that employed in the balanced modulator, namely, by applying the

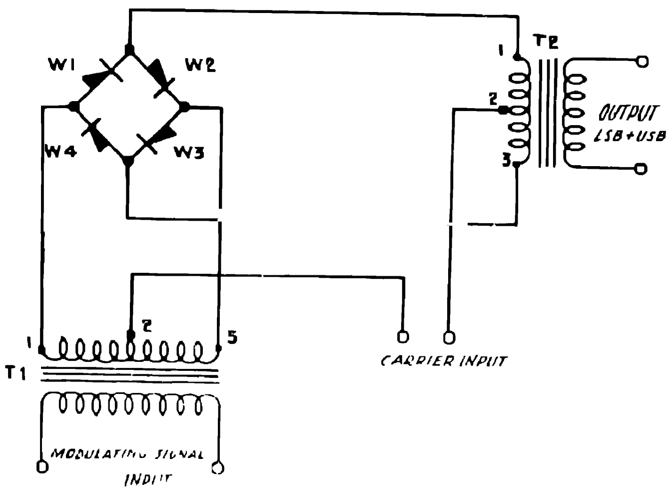


FIG. 503.—Double-balanced bridge-ring modulator.

carrier to the centre-taps of the input and output transformers, a carrier leak potentiometer being incorporated if required. In addition, the modulating signal is suppressed by arranging the four modulating rectifiers in the form of a bridge. The modulating

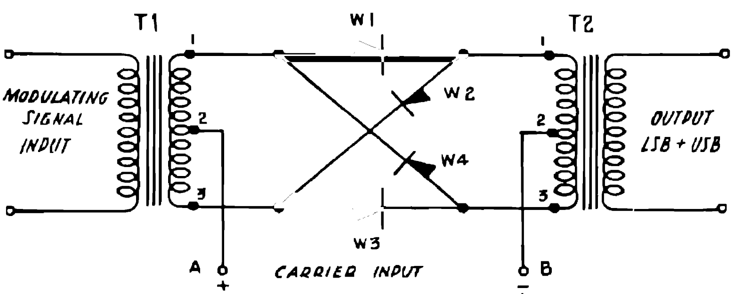


FIG. 504.—Fig. 503 redrawn in the lattice form.

signal input is applied across one diagonal of this bridge, the output being taken from the other diagonal, as shown in Fig. 503. In this manner, sidebands will be obtained in the output free from modulating signal or carrier frequencies.

This bridge network shown in Fig. 503 may be redrawn in the lattice form, as in Fig. 504.

To deduce the output waveform from such a modulator when the amplitude of the carrier input is large compared with that of the modulating signal, consider the biasing effect of the positive and negative half-cycles of the carrier input on the rectifiers.

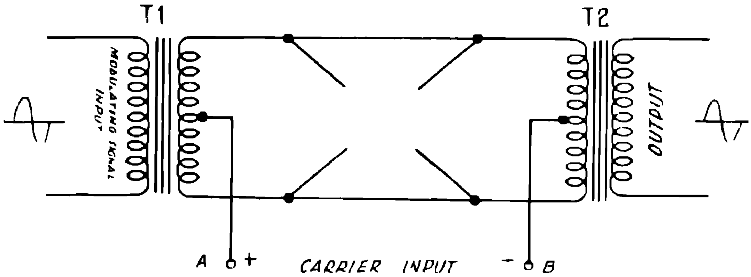


FIG. 505.—Double-balanced bridge-ring modulator with series rectifier elements forward-biased and parallel rectifier elements back-biased.

When A is at a higher potential than B, rectifiers $W1$ and $W3$, being forward-biased, will offer a low impedance path, whereas rectifiers $W2$ and $W4$, being back-biased, will offer only a high impedance path to the signal passing from transformer $T1$ to transformer $T2$. The majority of the modulating signal current will therefore flow through the low impedance offered by $W1$ and $W3$ (*see* Fig. 505).

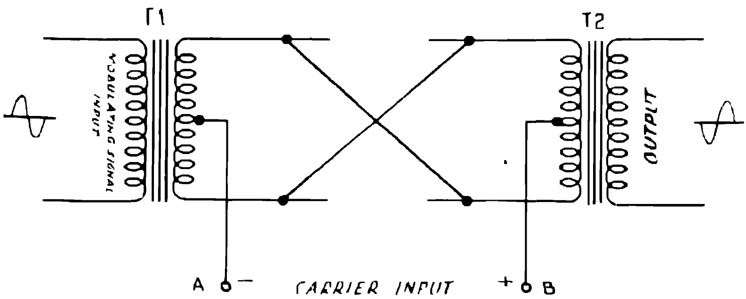


FIG. 506.—Double-balanced bridge-ring modulator with series rectifier elements back-biased and parallel rectifier elements forward-biased.

On the next half-cycle of carrier, B will be at a higher potential than A; rectifiers $W1$ and $W3$, now being back-biased, offer a high-impedance path, but rectifiers $W2$ and $W4$ provide a low impedance

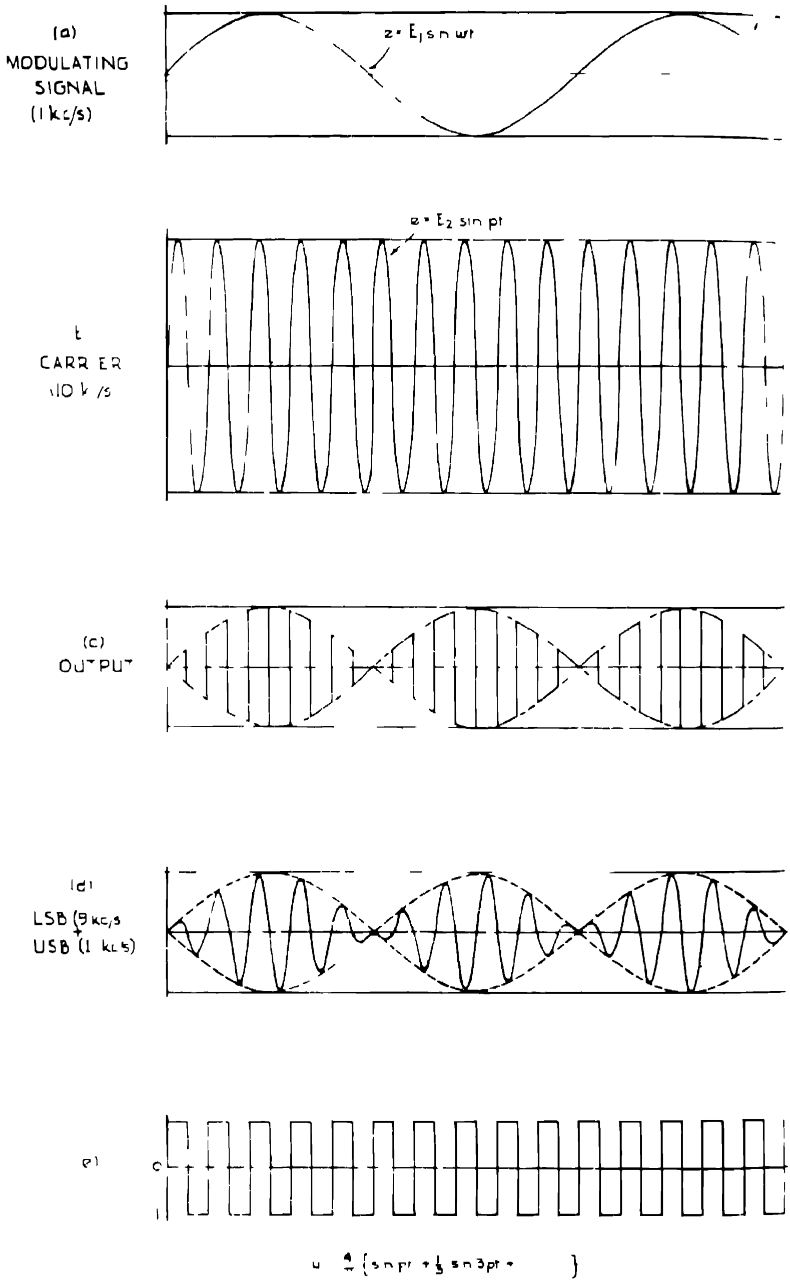


FIG. 507 —Waveform produced by double balanced modulator when the carrier frequency is ten times the modulating frequency

path. The majority of the signal will therefore pass from transformer $T1$ to $T2$ *via* the low impedance offered by $W2$ and $W4$ (*see* Fig. 506).

It will be noted that after each half-cycle of the carrier the signal in the output undergoes a phase-shift of 180°. Since the carrier frequency will, in general, be greater than that of the modulating signal, this change in phase will occur several times during each cycle of the input signal; the output waveform when the carrier frequency is ten times the modulating signal frequency is shown in Fig. 507*c*.

It will be noted that the output waveform *c* differs slightly from the beat note waveform *d* of the two sidebands. This is due to the presence of harmonics.

Inspection of the output waveform shows it to be equivalent to the modulating signal waveform multiplied by the square waveform *e*. Since the square waveform has unit amplitude, and the same frequency as the carrier, it may be represented by the equation :—

$$y = \frac{4}{\pi}\left[\sin pt + \tfrac{1}{3}\sin 3\,pt + \tfrac{1}{5}\sin 5\,pt + \ldots\right]$$

Thus the output waveform may be represented by :—

$$E_1 \sin \omega t \times \frac{4}{\pi}\left[\sin pt + \tfrac{1}{3}\sin 3\,pt + \tfrac{1}{5}\sin 5\,pt + \ldots\right]$$

$$= \frac{4E_1}{\pi}\left[\sin \omega t \sin pt + \tfrac{1}{3}\sin \omega t \sin 3\,pt + \tfrac{1}{5}\sin \omega t \sin 5\,pt + \ldots\right]$$

$$= \frac{2E_1}{\pi}\left[\cos (p - \omega)t - \cos (p + \omega)t\right]$$

$$+ \frac{2E_1}{3\pi}\left[\cos (3p - \omega)t - \cos (3p + \omega)t\right]$$

$$+ \frac{2E_1}{5\pi}\left[\cos (5p - \omega)t - \cos (5p + \omega)t\right] + \ldots$$

It will be seen that this result is similar to that obtained for the balanced modulator. However, the modulating signal component is no longer present, and by comparing the equations it will be seen that the sidebands have twice the amplitude, indicating that an improvement in the output is obtained by the use of the double-balanced bridge-ring modulator.

Other forms of double-balanced bridge-ring modulators

Several arrangements of the double-balanced principle are now used. Fig. 508 shows a double-balanced modulator having an improved rectifier biasing system.

This circuit may be redrawn to emphasise the double-balanced structure by considering $T2$ as two transformers in series (Fig. 509).

To explain its operation, Fig. 508 is redrawn once again as in Fig. 510.

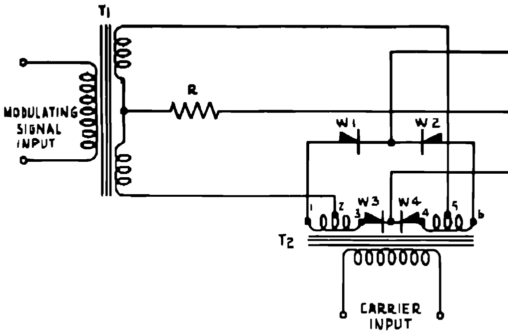


FIG. 508.—Double-balanced modulator—alternative form.

Considering the positive half-cycle of the carrier, rectifiers $W1$ and $W4$ are forward-biased, rectifiers $W2$ and $W3$ are back-biased (Fig. 511*a*). Considering the negative half-cycle of the carrier, rectifiers $W2$ and $W3$ are forward-biased, rectifiers $W1$ and $W4$ are back-biased (Fig. 511*b*).

It follows that the input signal undergoes a phase reversal on every half-cycle of the carrier, and the output waveform must therefore be the same as that produced by the first bridge-ring modulator discussed (*see* Fig. 507*c*). The output thus contains the two sidebands, but no carrier frequency and no modulating signal frequency components.

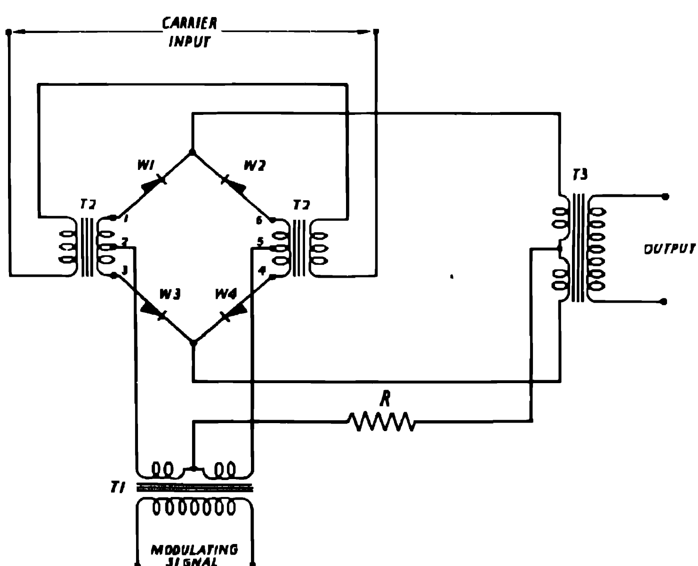


FIG. 509.—Double-balanced modulator (redrawn to explain operation).

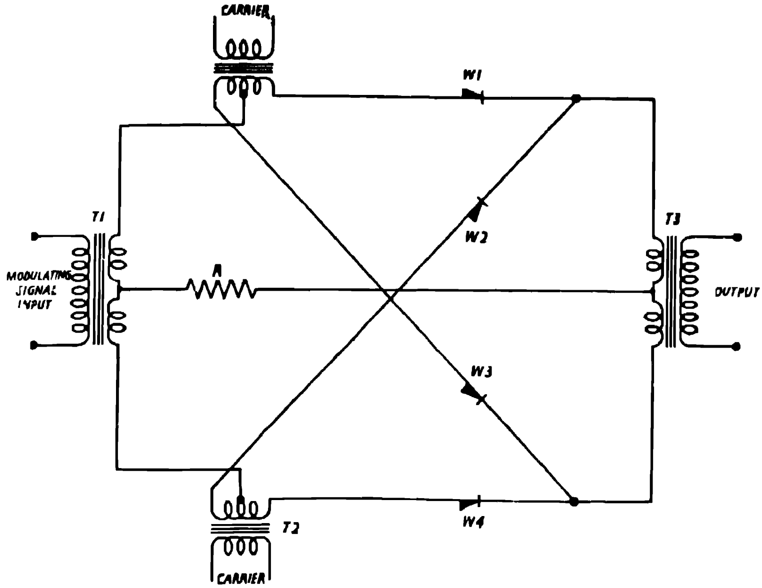


FIG. 510.—Double-balanced modulator (redrawn to explain operation).

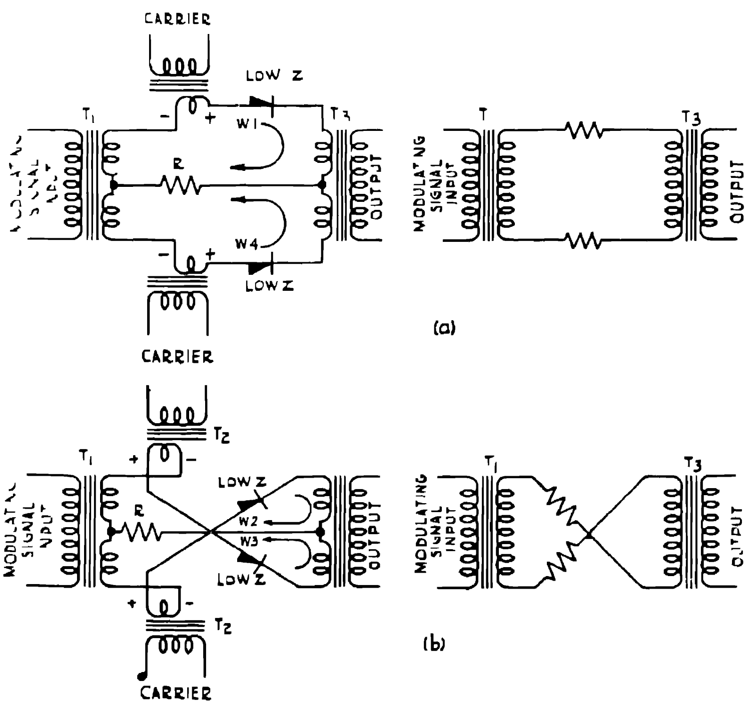


FIG. 511.—Operation of double-balanced modulator.

It will be noted that R limits the carrier current flowing, and that the potential drop across R tends to back-bias all the rectifiers. However, when in operation, one pair of rectifiers is forward-biased by half the windings of $T2$ whilst the other pair are back-biased by the remaining windings.

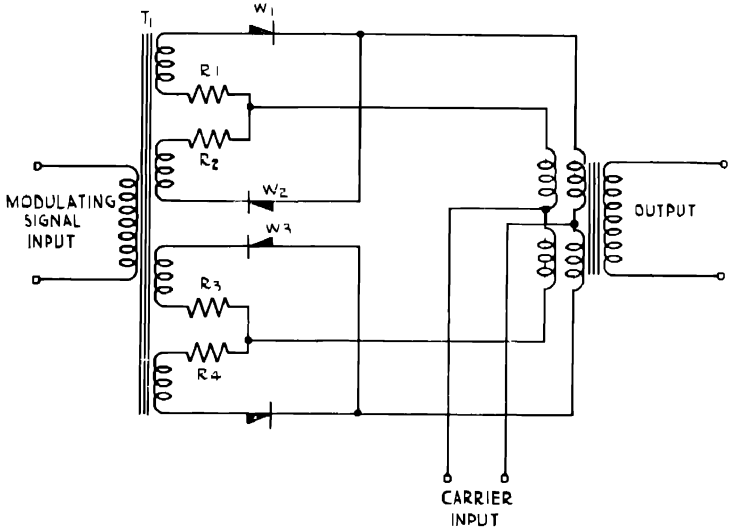


FIG. 512.—Alternative type of double-balanced modulator.

Fig. 512 shows an alternative type of modulator in use; the principle of operation is very similar and it gives the same output waveform (Fig. 507c).

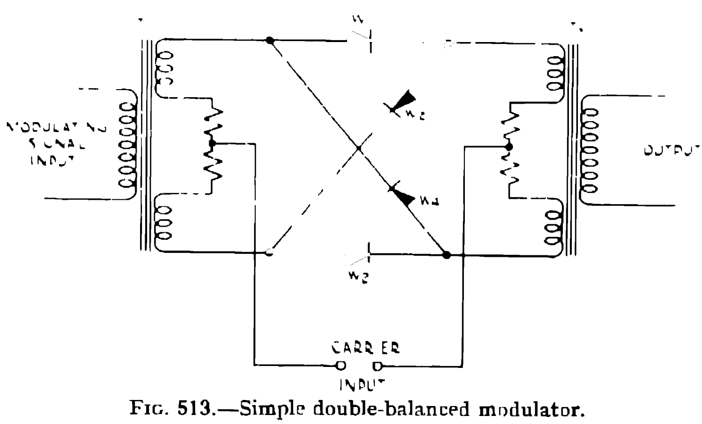


FIG. 513.—Simple double-balanced modulator.

Its derivation from the simple double-balanced modulator may be seen from Figs. 513, 514 and 515. Fig. 515 will be seen to be identical with Fig. 512.

Replacing $T1$ by a transformer having one primary and four secondary windings, Fig. 514 is obtained.

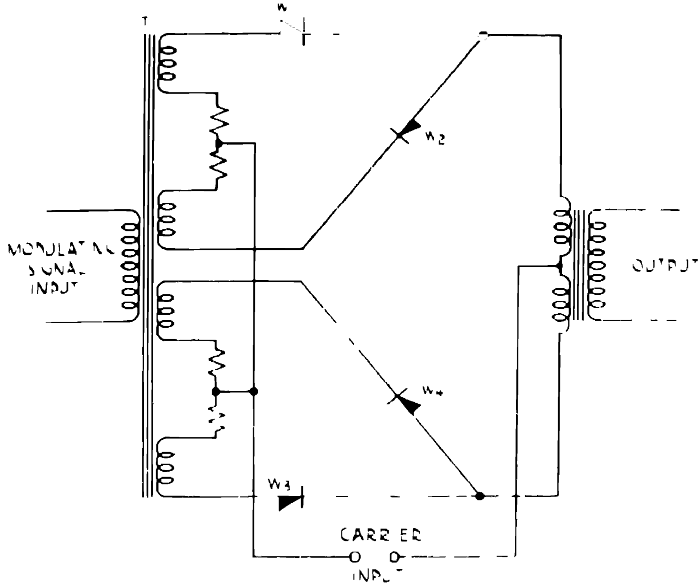


Fig. 514.—Derivation of alternative type of double-balanced modulator.

Replacing $T2$ by a five-winding transformer gives Fig. 515.

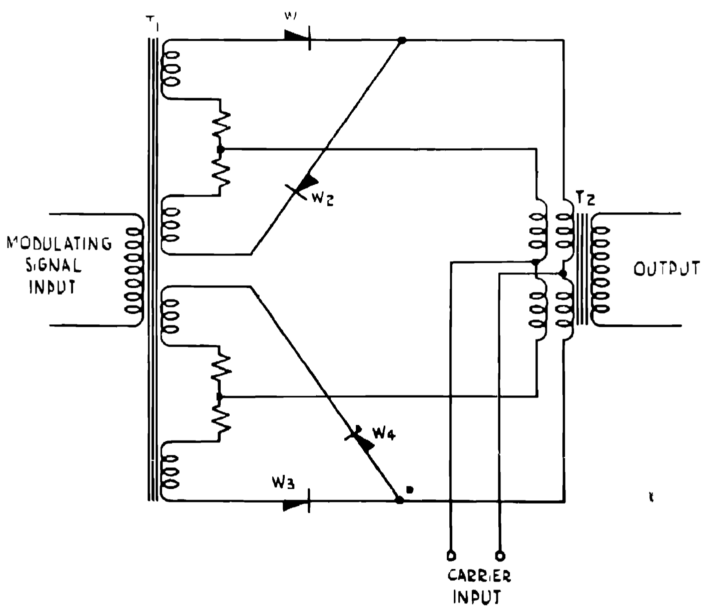


Fig. 515.—Alternative type of double-balanced modulator.

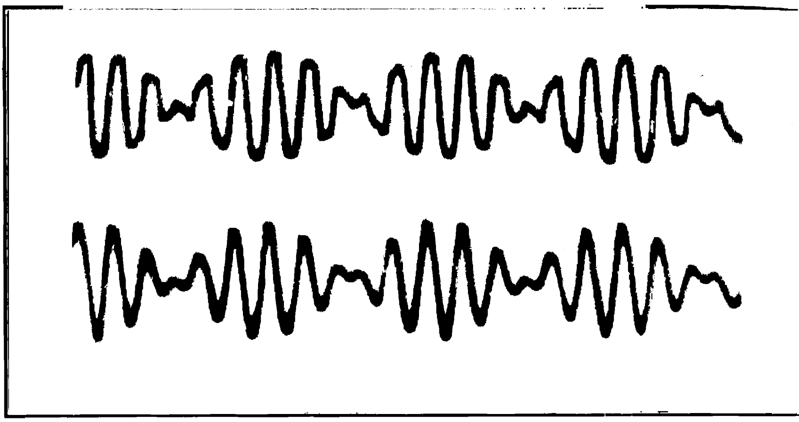


PLATE 28.—CRO traces using a double-beam tube.

(a) Output of bridge-ring modulator (above).
(b) Addition waveform of two sine waves having the same frequencies as upper and lower sidebands (below).

DEMODULATORS

The function of a demodulator is to reproduce the modulating frequency. The method adopted in the process of demodulation depends on whether or not the carrier frequency is present in the received waveform. If present, demodulation may be carried out simply by using a non-linear impedance, this process being known as detection.

Linear detection

When the non-linear impedance adopted is the metal rectifier or the diode valve, the process of detection resolves itself into one of rectification (*see* Fig. 516).

Assuming the non-linear impedance employed has a characteristic as shown in Fig. 517, the waveform after rectification can be considered to be the product of the input waveform and a square waveform of carrier frequency (Fig. 518).

The output waveform

$$= A \sin pt\left(1 + \frac{a}{A}\sin \omega t\right)\left[\frac{1}{2} + \frac{2}{\pi}\left(\sin pt + \frac{1}{3}\sin 3\,pt + \ldots\right]\right.$$

$$= \ldots + \frac{2}{\pi}\,.\,\sin pt\,.\,A \sin pt\left(1 + \frac{a}{A}\sin \omega t\right)\ldots$$

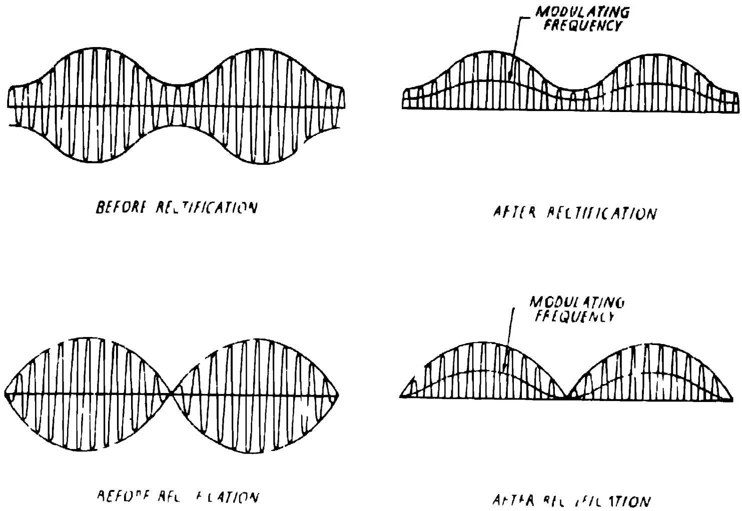


FIG. 516 — *Above* Carrier plus both sidebands
Below Carrier plus one sideband (either USB or LSB)

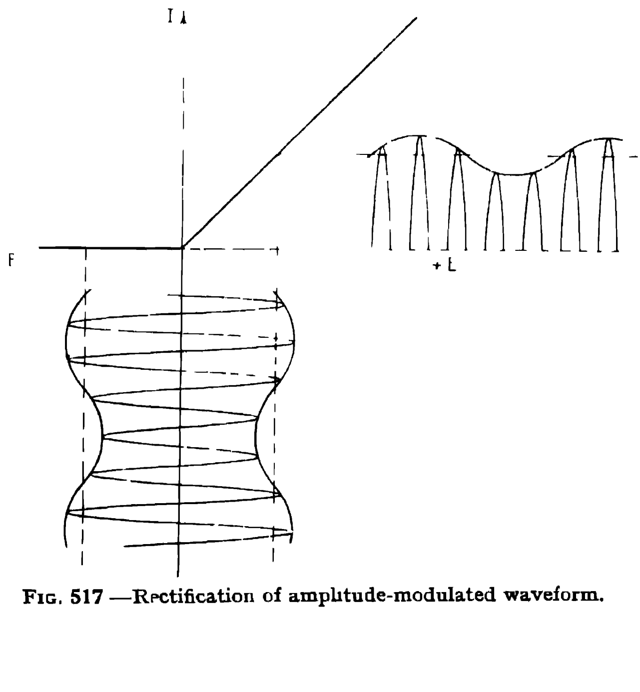


FIG. 517 —Rectification of amplitude-modulated waveform.

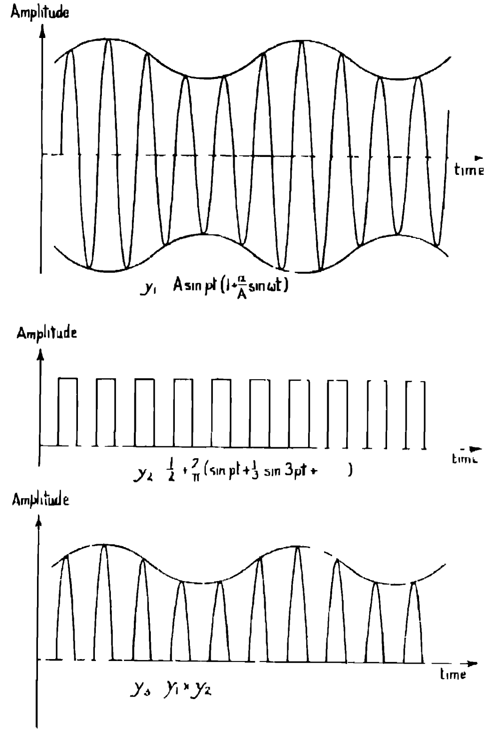


FIG 518—Rectifier output considered as the product of the modulated carrier input, and a square wave of carrier frequency.

$$= \ldots + \frac{A}{\pi}\left(1 - \cos 2pt\right)\left(1 + \frac{a}{A}\sin \omega t\right)\ldots$$

$$= \ldots + \underset{\substack{\uparrow \\ \text{DC}}}{\frac{A}{\pi}} - \frac{A}{\pi}\cos 2pt + \underset{\substack{\uparrow \\ \text{Modulating frequency}}}{\frac{a}{\pi} \,.\, \sin \omega t} - \frac{a}{\pi} \,.\, \sin \omega t \,.\, \cos 2pt$$

The term $\frac{a}{\pi}$ sin ωt corresponds to the original modulating frequency, which is the lowest alternating frequency in the output. It may be separated from all the other components of detection by means of either a low-pass filter or a resistance-capacity network.

The diode as a detector

A diode valve may be used as a detector (or demodulator) for an amplitude modulated signal When the carrier frequency is high compared with the modulating frequency, a resistance-capacity network may be used to separate out the required modulating frequency from the products of detection. Such circuits will now be considered.

When used as a detector, the diode may be arranged in the form of either a series circuit (Fig. 519*a*) or a shunt circuit (Fig. 519*b*). In each case, *R* is the load resistor across which potentials at the modulating frequency are required. Consider first the series diode of Fig. 519*a*. If an unmodulated carrier signal were applied to the input of the circuit and the load resistance *R* were disconnected from the circuit, electrons would flow in the direction shown, whenever the anode were made positive with respect to the cathode —that is, on every "positive" half-cycle—until the potential across *C* were equal to the peak voltage of the incoming carrier signal. The presence of the load resistor *R* causes *C* to discharge slightly during the half-cycle that makes the anode negative with respect to the cathode. Thus for an unmodulated input a voltage

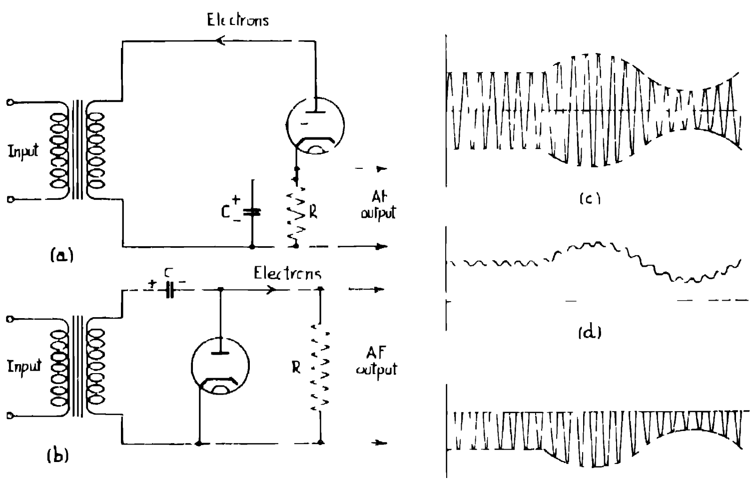


FIG. 519.—The diode as a detector, showing (*a*) series, and (*b*) shunt connections.

is developed across *R* which is composed of a DC voltage slightly smaller than the peak voltage of the carrier, and a carrier frequency ripple whose magnitude will depend on the time constant of *R* and *C*. If the time constant is large, the voltage across *C* will drop very little during one half-cycle of the carrier, and the ripple will be small. If the applied signal is a modulated carrier signal, as in Fig. 519*c*, and the time constant of *C* and *R*, although large compared with the periodic time of the carrier, is small compared with the periodic time of the highest modulating frequency likely to be encountered, then the voltage developed across *R* will have the form of Fig. 519*d*. That is to say, the voltage across *R* will still be slightly less than the peak voltage of the applied carrier (together with a carrier ripple), but in view of the low value of time constant of *C* and *R* (compared with the periodic time of the audio

modulating frequency), this voltage across R will be able to follow the audio-frequency variations in the applied HF peak voltage. The alternating component of the voltage across R is then fed *via* a DC blocking condenser to the grid of the next stage, which will be an audio amplifier.

The shunt diode (Fig. 519*b*) will give similar results. On the half-cycle of carrier that makes the anode positive with respect to cathode, electrons will flow from cathode to anode, charging the condenser C with polarity as shown. On the other half-cycle, electrons will flow downwards through the load resistance R. Across C is developed a DC voltage, plus a carrier ripple, and also an audio-frequency variation if the input is modulated and the time constant of C and R is of suitable value. The voltage across R will be as shown in Fig. 519*e*.

Metal rectifier as detector

In the above example of diode detection the diode valve may be replaced by a metal rectifier. In fact any method of rectification employing metal rectifiers may be used as a form of detector.

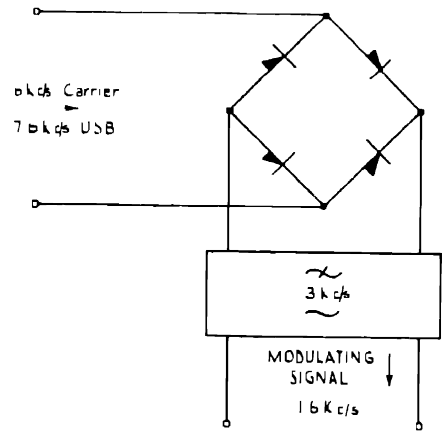


FIG. 520.—Metal rectifier detector.

Fig. 520 shows the use of a full-wave bridge rectifier as a demodulator, a low-pass filter being employed, in this case, to eliminate the unwanted products of demodulation. To ensure efficient operation, a path must be provided, usually within the LP filter, for the DC component of demodulation.

Square-law detection

Square-law detection is the name given to detection by means of a device such that the application of a sinusoidal input gives rise to components in the output proportional to the square of the input. Anode bend detection employing a valve biased to

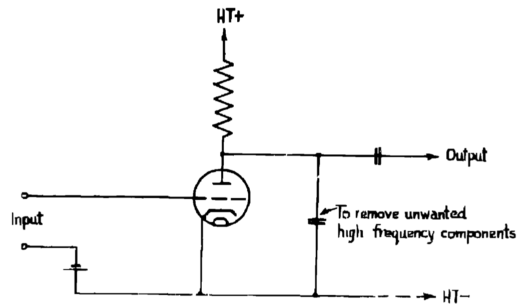


FIG. 521.—Anode bend detector.

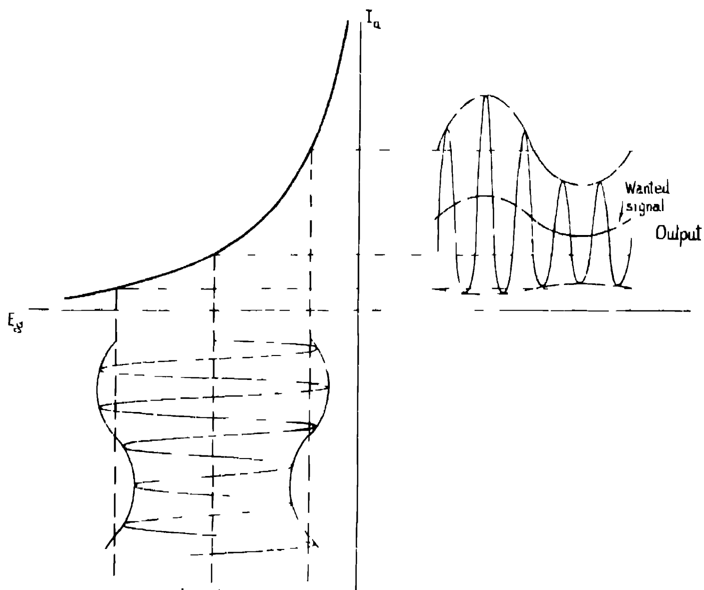


FIG. 522.—Principle of square-law (anode bend) detection.

operate on the curved portion of its characteristic is an example of square-law detection (*see* Fig. 521 and Fig. 522).

A summary of the principal frequencies produced by both square-law modulation and square-law detection are given in Fig. 523.

f = frequency of modulating signal.

f_c = frequency of carrier.

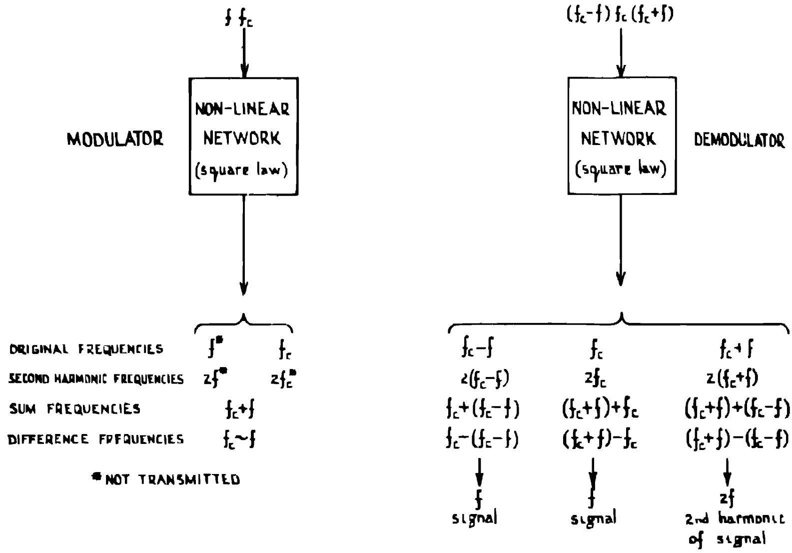


FIG. 523 – Frequencies produced by square-law modulation and demodulation.

DEMODULATION IN SUPPRESSED CARRIER SYSTEMS

As has already been stated, when only a single sideband and no carrier is transmitted, the modulating frequency may be reproduced by *modulating* the received waveform, using a local oscillator to

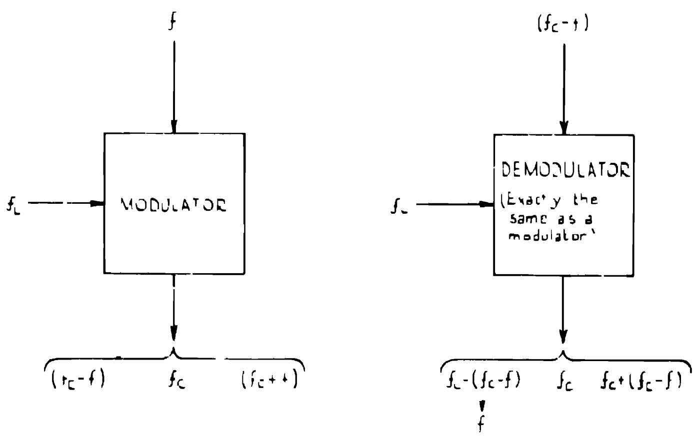


FIG. 524.—Frequencies produced by modulation and demodulation.

generate the carrier frequency. Any of the forms of modulator described above may be employed. Suppose that the lower sideband $(f_c - f)$ is received: modulating it with the carrier f_c will produce an upper sideband $(2f_c - f)$ and a lower sideband f (*see* Fig. 524). Since only the lower sideband is required, the upper sideband must be removed by means of a filter.

Demodulation using simple balanced modulator

The circuit used for modulation (*see* Fig. 497) may also be used for demodulation. The carrier frequency generated by the local oscillator is fed into the centre taps of the transformers, as before, the incoming sideband being applied to transformer T_1. The output

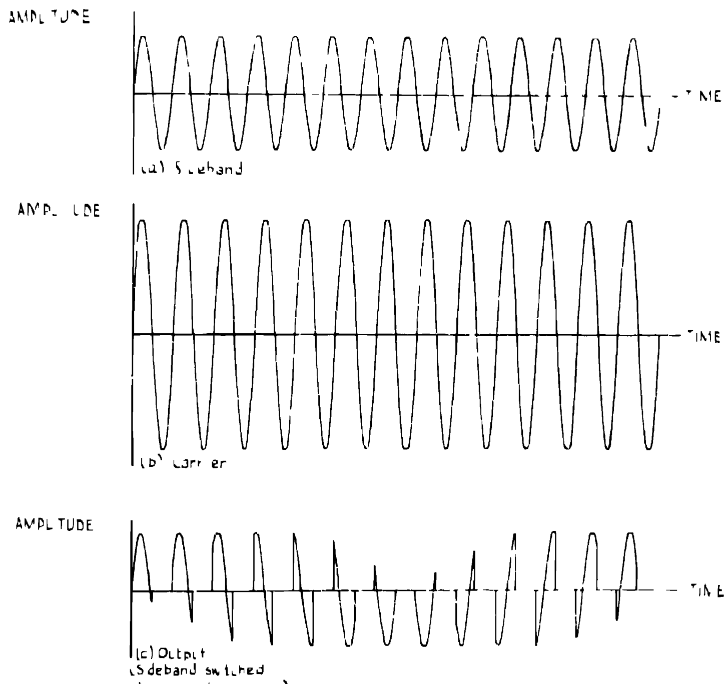


FIG. 525.—Simple balanced demodulator waveforms.

waveform (*see* Fig. 525*c*) is produced by the sideband being switched on and off at the carrier frequency. This output waveform contains a component corresponding to the original modulating signal. This may be shown mathematically :—

Let the received sideband be :—

$$e_1 = E_1 \sin (p + \omega)t$$

and let the local oscillator generate a signal :—

$$e_2 = E_2 \sin (p + \delta)t,$$

which represents a signal having a slightly different frequency from that originally fed to the modulator at the sending end.

If E_2 is large compared with E_1, then the output waveform corresponds to a square waveform of unit amplitude and frequency $\frac{p+\delta}{2\pi}$, multiplied by the received sideband (*see* pages 508–9).

The output waveform y_3 is therefore :—

$$y_3 = \frac{E_1}{2}.\sin(p+\omega)t + \frac{2E_1}{\pi}\sin(p+\omega)t\,.\left\{\sin(p+\delta)t + \frac{1}{3}\sin 3(p+\delta)t + \ldots\right\}$$

$$= \frac{E_1}{2}\sin(p+\omega)t + \frac{2E_1}{\pi}\left\{\sin(p+\omega)t\,.\sin(p+\delta)t + \frac{1}{3}\sin(p+\omega)t\,.\sin 3(p+\delta)t + \ldots\right\}$$

$$= \frac{E_1}{2}.\sin(p+\omega)t + \frac{E_1}{\pi}\left\{\cos(\omega-\delta)t - \cos(2p+\omega+\delta)t\right\}$$
$$+ \frac{E_1}{3\pi}\left\{\cos(2p-\omega+3\delta)t - \cos(4p+\omega+3\delta)t\right\} + \ldots$$

The required signal is represented by the term $\frac{E_1}{\pi}\cos(\omega-\delta)t$, and it will be seen that its frequency has been changed by $\frac{\delta}{2\pi}$. This shows that the signal obtained will have the same frequency as the modulating signal only when the two carrier oscillators are exactly synchronised, otherwise the received frequency will have been changed by the amount of the discrepancy between the two oscillators

Provided only one sideband has been transmitted, the phase with which the carrier frequency is replaced is of no consequence.

Demodulation using double-balanced modulator

In a similar manner, the double balanced modulators shown in Figs. 503, 508 and 512 may be used for demodulation Once again, the carrier frequency must be generated by a local oscillator.

The output waveform (*see* Fig 526*c*) is produced by introducing 180° phase-shifts in the sideband at the carrier frequency. Owing to the double-balanced nature of the modulator, there will be neither sideband nor carrier components in the output The output will, however, contain a component corresponding to the original modulating signal

This may be shown mathematically; assume that the received sideband is :—

$$y_1 = E_1 \sin(p+\omega)\,.\,t$$

and that the local oscillator generates a signal

$$y_2 = E_2\,.\sin(p+\delta)\,.\,t$$

then the output waveform will be :—

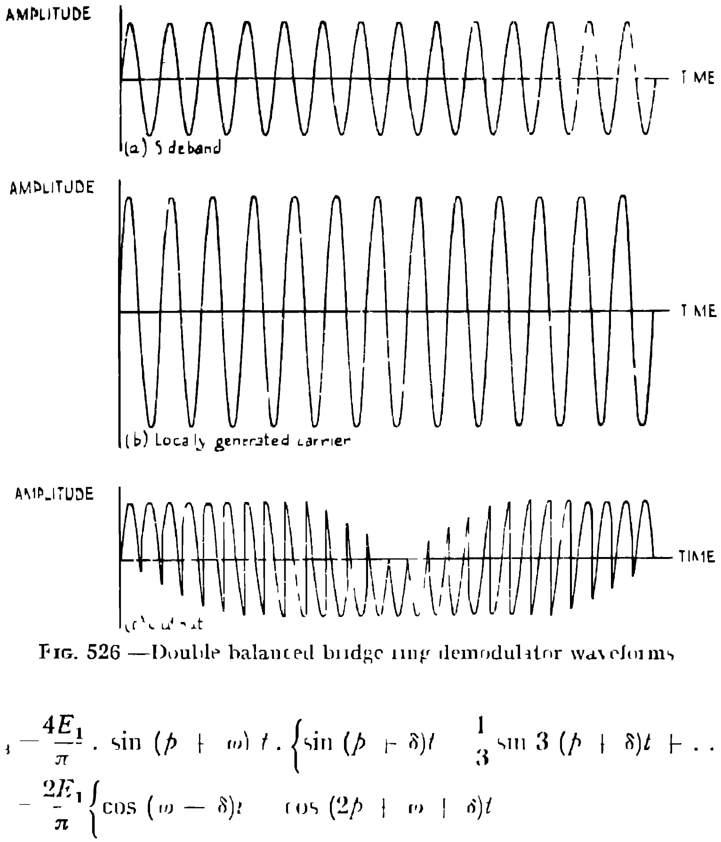


FIG. 526 —Double balanced bridge ring demodulator waveforms

$$v_3 = \frac{4E_1}{\pi} . \sin(p+\omega)t . \left\{\sin(p+\delta)t + \frac{1}{3}\sin 3(p+\delta)t + \ldots\right\}$$

$$= \frac{2E_1}{\pi}\left\{\cos(\omega-\delta)t - \cos(2p+\omega+\delta)t \right.$$

$$\left. + \frac{1}{3}\cos(2p-\omega+3\delta)t - \frac{1}{3}\cos(4p+\omega+3\delta)t + \ldots\right\}$$

The required signal component is $\frac{2E_1}{\pi}\cos(\omega-\delta)t$, *i.e.*, it has twice the amplitude of that obtained in the case of the simple balanced modulator. The same need for synchronism between the two carrier oscillators is still applicable.

AMPLITUDE, PHASE, AND FREQUENCY MODULATION

So far mention has only been made of amplitude modulation. There are two other modulation systems occasionally used in line communication, namely, "phase modulation" and "frequency modulation". Briefly, the characteristic features of the three systems are as follows :—

(1) *Amplitude modulation* (*AM*).—The intelligence is contained in the amplitude term ; the carrier amplitude is usually a linear function of the instantaneous amplitude of the modulating waveform.

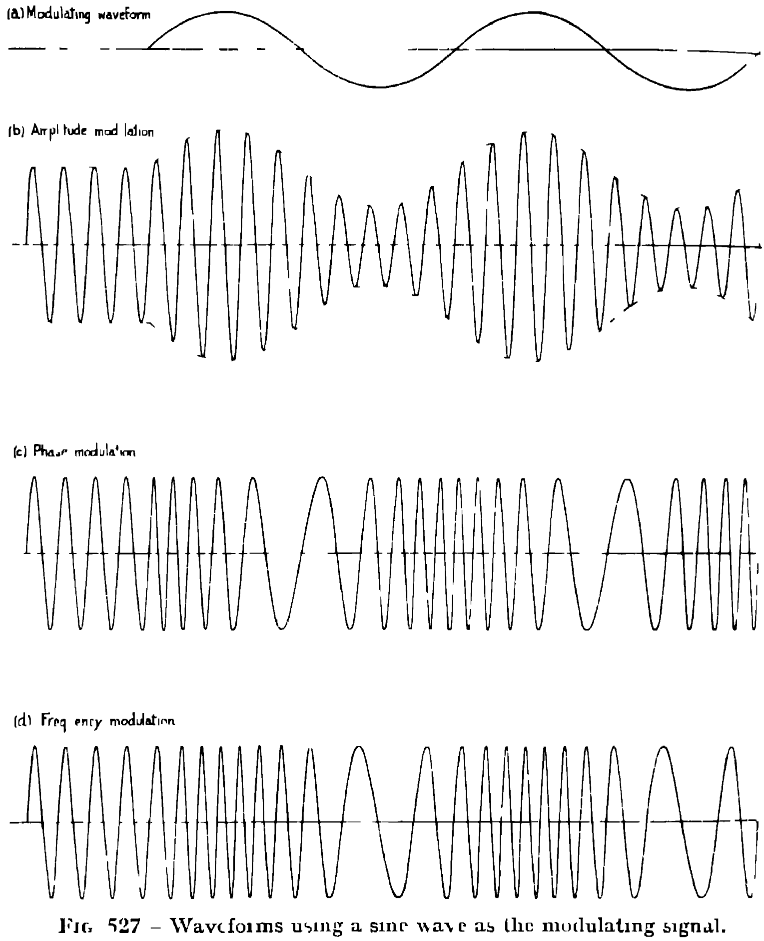


FIG. 527 – Waveforms using a sine wave as the modulating signal.

(2) *Phase modulation* (*PM*) —The amplitude of the carrier is constant; the intelligence is imposed by causing the instantaneous phase term to be a linear function of the instantaneous amplitude of the modulating waveform.

(3) *Frequency modulation* (*FM*).—The instantaneous frequency is a linear function of the instantaneous amplitude of the modulating waveform

Fig. 527 shows a carrier (*b*) amplitude-, (*c*) phase-, and (*d*) frequency-modulated by a given sinusoidal modulating signal.

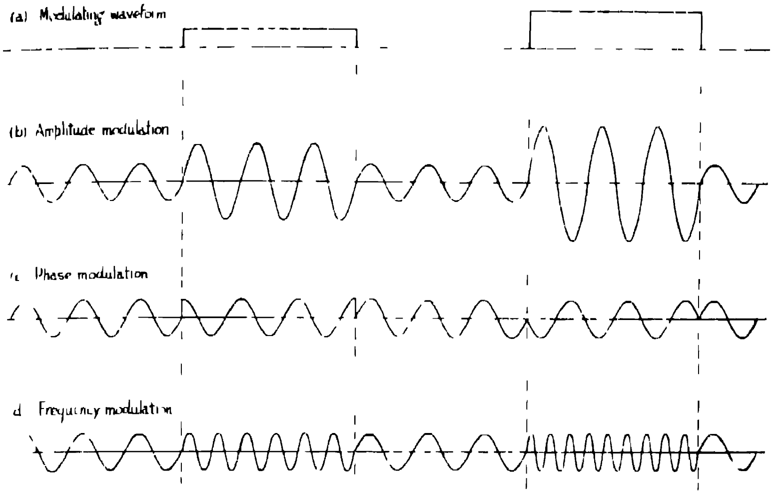


FIG. 528.—Waveforms using a square-wave as the modulating signal.

Fig. 528 shows a carrier (*b*) amplitude-, (*c*) phase-, and (*d*) frequency-modulated by a square wave modulating signal.

CHAPTER 12

THE CATHODE RAY OSCILLOSCOPE

The study and measurement of alternating currents and voltages is a most important one from the communications point of view It has been shown that meters may be constructed which give certain restricted information about, say, an alternating voltage Such a voltmeter records either the RMS, the mean, or the peak value of the voltage, depending on the calibration, but it gives no indication of the purity of the waveform; that is to say, a meter cannot differentiate between a voltage having components at different frequencies and a pure sinusoidal voltage of the same RMS (or mean or peak) value. Nor, if used to measure voltages at different parts of a circuit, can a meter give information regarding phase relationship.

A meter suffers also from other disadvantages due to the inertia of its moving parts. For instance, the pointer does not deflect instantaneously, and having been deflected, does not immediately come to rest, however " dead beat " the action may be. All voltmeters have a finite impedance and represent a load on any circuit to which they are applied, taking power from it and thereby disturbing the circuit conditions and giving rise to a false reading In addition, no meter is completely satisfactory over a large range of frequencies such as is encountered in communication engineering

The cathode ray oscilloscope (or " CRO ", as it is usually called) overcomes all these disadvantages, and gives a complete graphical representation of an alternating quantity. It is thus a valuable instrument for use in communication engineering.

The principle is that a stream of electrons is focused into a narrow beam and made to impinge on a fluorescent screen; this glows visibly where hit by the electron beam, which thus produces a spot of light. The electron beam, and the spot it produces on the screen, can be deflected horizontally and vertically by two independent deflector systems, the deflection produced being proportional to the voltage (or current) applied to the system.

If, for example, it is desired to examine the waveform of a 50 c/s voltage, the spot is made to move at a uniform speed across the screen from left to right, and at the same time the alternating voltage is applied to give a vertical deflection. Since the horizontal deflection at any instant is proportional to time, and the vertical deflection is proportional to the instantaneous value of the voltage,

the spot will trace out a graph showing the instantaneous voltage as a function of time. When the spot reaches the right-hand side of the screen it can be brought back very rapidly, and in such a way that when the process is repeated the spot traces out the same path as on the previous occasion. In this way a stationary picture of the waveform is obtained

The construction of the tube and the principles underlying its operation will now be discussed in greater detail.

THE CATHODE RAY TUBE

A simple cathode ray tube is shown in Fig. 529. It consists of an evacuated glass bulb containing a heated cathode, which may be directly or indirectly heated, and an anode which is maintained at a positive potential relative to the cathode. The anode is made

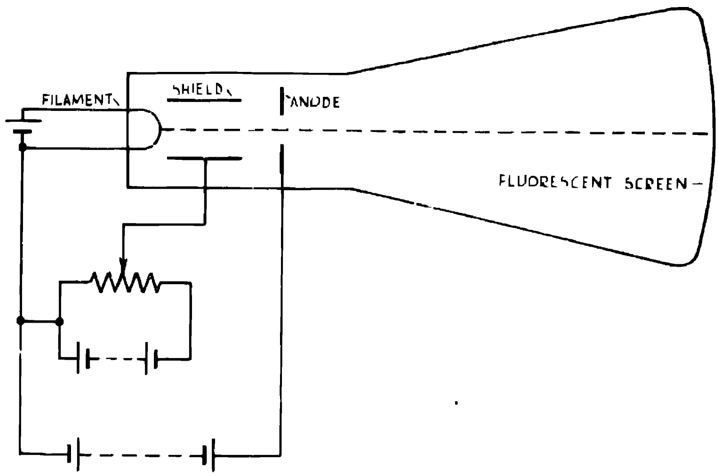


FIG. 529.—Simple cathode ray tube

in the form of a disc having a hole in the centre, and is situated close to the cathode. Electrons emitted by the cathode are attracted towards the anode. Most of the electrons will be attracted to the anode, but some will pass through the central hole and continue to travel in a straight line until impinging on the fluorescent screen at the end of the tube. This screen is formed by coating the inside of the glass with a fluorescent substance which glows under the impact of electrons. Where the electrons strike this fluorescent screen, therefore, a patch of light appears.

In order to increase the number of electrons passing through the central anode aperture, a cylindrical shield is fitted so as to concentrate the electrons in transit from cathode to anode into a narrow beam. This effect is obtained by maintaining the shield at a negative potential relative to the cathode. The electrons are repelled by this shield, and the electron stream is therefore concentrated along its axis, thereby increasing the proportion of electrons that pass through the anode.

Brilliancy

The brilliancy of the light spot on the fluorescent screen will depend on the energy contained in the electron stream—that is on the number and velocity of the electrons bombarding the screen at any instant. In order to obtain a sufficiently powerful electron stream, an anode potential of the order of 1000 volts is required in the normal hot-cathode or low voltage tube. The potential of the shield will to a certain extent affect the electron stream in the same way that the grid potential affects the anode current in a thermionic valve.

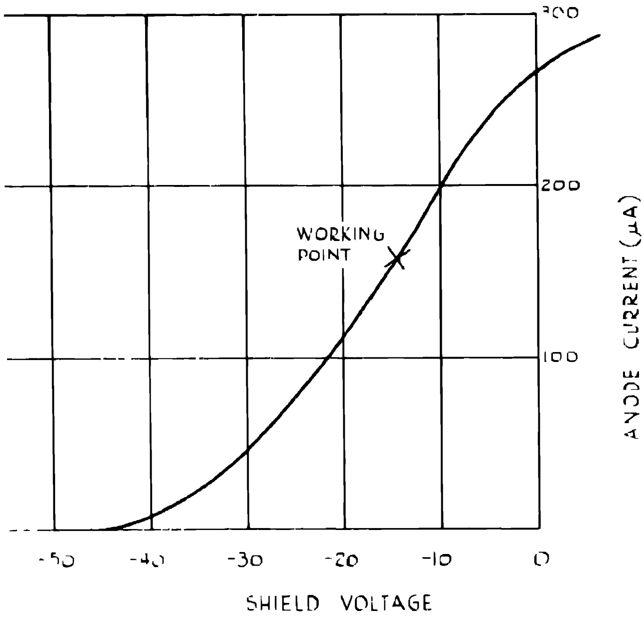


FIG. 530.—Characteristic of cathode ray tube.

Fig. 530 shows how the shield potential affects the anode current ; it will be seen that the more negative the shield becomes the smaller the number of electrons drawn towards the anode but, due to the concentrating effect of the shield, the greater the percentage of these electrons arriving on the fluorescent screen These two factors have a conflicting effect on the number of electrons actually arriving at the screen, and at some point on the curve the screen will receive the maximum number of electrons and the light spot will have maximum brilliance.

The fluorescent screen

Many substances are fluorescent ; that is, they have the property that they emit light when subjected to electron bombardment. This property is, however, possessed by different substance

to different degrees ; for instance, one substance may emit a greater intensity of light than another for a given rate of bombardment ; again, the frequency—that is, the colour of the light emitted—varies from substance to substance. All fluorescent materials continue to emit light for some time after the electron bombardment has ceased ; this is known as " after-glow ", and its duration varies with different substances from a few microseconds to many seconds. In some applications, it is an advantage for a tube to have a long after-glow.

The phenomenon of fluorescence has not yet been satisfactorily explained, but certain general observations may be made. In general, fluorescent materials consist of a crystalline metallic salt containing a minute trace of impurity. This impurity, known as the activator, is essential, since the base substance in its pure state frequently has no fluorescent properties. The base and the activator together determine the fluorescent properties described above. A short representative list of fluorescent materials used for the screens of cathode ray tubes is given below.

TABLE XVI
Fluorescent screen materials

Base	*Activator*	*Colour of trace*
*Zinc silicate	manganese	blue-green
Zinc sulphide	manganese	orange
Zinc sulphide	silver	blue
Zinc sulphide	copper	green

* Occurs in nature as " willemite ", and is the material most commonly used for CRO screens.

In selecting a material for the screen of a cathode ray tube, these various properties must be considered in relation to the use to which the tube is to be applied. A screen that emits a high intensity of light for a given rate of electron bombardment is desirable in practically all cases, since otherwise high anode voltages would be required to produce the required brilliancy of the image. It is also essential that the substance can be applied to the end of the tube in such a way that it produces a uniform screen.

If the tube is to be used for visual examination of waveforms, as is usually the case, the trace must be of a colour that produces minimum fatigue and eye strain whether viewed in daylight or in artificial lighting. The best colour for this purpose is found to be green, though blue may be slightly better in artificial light. If the waveform under examination is recurrent, and the spot of light may be made to trace the same path again and again, an after-glow of 10 or 20 microseconds will be sufficient, with the natural persistence of vision, to give the impression of a stationary trace

at all but the lowest frequencies. For visual examination of very low-frequency waveforms, and in particular for transients, *i.e.*, non-recurrent waveforms, a longer after-glow is desirable and may be of the order of several seconds.

If the tube is intended only for photographic work a blue trace is desirable, the blue light being more "actinic" than the green, that is, it has a greater effect on the light-sensitive material of the photographic film for a given exposure. In a general purpose tube intended for both photographic and visual examination, a screen is used that gives a blue-green trace. Tubes for television purposes are required to give a white trace, which is obtained by using a mixture of two or more fluorescent materials each giving a different portion of the spectrum.

FOCUSING

For oscillograph work, a very small sharply defined light spot is required; and although a certain amount of focusing can be obtained by adjustment of the shield potential, this in itself is not sufficient, and it is necessary to adopt some additional focusing device. There are three principal methods by which this may be done. The first is by using a "soft" or gas-focused tube, which has an inherent focusing effect on the electron stream; the second method is electrostatic focusing, in which the electron stream is passed through an electrostatic field that is so shaped as to cause the electrons to converge on the screen. This type of focusing has the advantage that it can be controlled easily. The third method electromagnetic focusing, is rarely used, except in tubes intended for television work.

Gas-focusing

In the gas-focused or soft tube, a high vacuum is first formed in the normal way, and then a small quantity of inert gas such as argon or helium is introduced. The passage of the electron beam through the rarified gas has the effect of ionising the gas—that is, splitting the atoms into free electrons and positive ions. The free electrons add themselves to the electron beam, and the relatively heavy positive ions drift slowly towards the cathode. Due to their mutual attraction the electrons and ions form a narrow beam, that is, the presence of the positive ions exerts an automatic and very effective focusing effect on the electron beam. This type of tube has two main disadvantages :—

(*a*) *Loss of focus at high frequencies.*—At high deflecting speeds, the relatively heavy positive ions tend to lag behind the electron beam, due to their greater inertia, resulting in loss of focus at high frequencies.

(*b*) *Limited life.*—The positive ions moving towards the cathode ultimately strike it, doing appreciable damage to the sensitive emitting surface. This limits the life of the tube t a few hundred hours.

Electrostatic focusing

The electrostatic method of focusing employs a complex anode system. Fig. 531 shows an arrangement using three anodes. This is the configuration of electrodes used nowadays in the majority of cathode ray tubes, though the potentials applied to the various

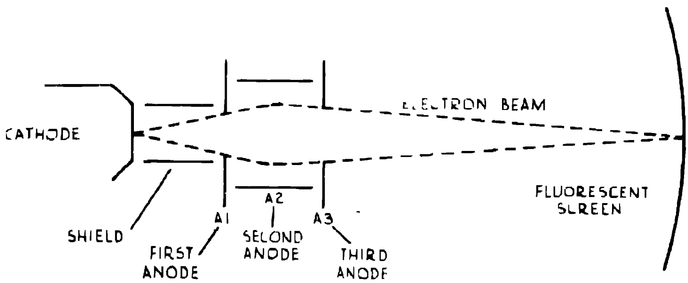


FIG. 531.—Electrostatic focusing in a three-anode tube.

anodes differ. The whole electrode system forms an electronic " lens " which acts on the electron beam in much the same way as does a converging optical lens on a beam of light. The theory of electron optics, which explains the action of an electric lens, will not be discussed; in any case the practical electric lens is usually designed empirically. In general, the first and third anodes are kept at fixed potentials and the shape of the electric field is controlled by varying the potential on the second anode; in this way the focal length of the electric lens may be adjusted to give optimum focusing of the electron beam, producing a very small spot of light on the fluorescent screen. Brilliancy is still controlled by the shield potential.

The potential of the shield, however, has a distinct effect on the focus, as has also the potential of the focusing anode on brilliancy. The result is that to a certain extent the two controls must be adjusted together to give optimum focus at the required brilliancy. It may be stated here that the trace should never be brighter than is absolutely necessary; the brighter the trace, the greater the impact of the electron beam on the fluorescent screen, and the shorter the life of the fluorescent material.

Electromagnetic focusing

This type of focusing was used in the earliest tubes and may still be found in the case of tubes used for particular purposes. It is not generally so convenient as electrostatic focusing for a general purpose tube.

In an electromagnetically focused tube, a simple electrode structure is required consisting simply of cathode, negative potential shield and high potential perforated anode; focusing is then carried out by means of a coil surrounding the neck of the tube, as

shown in Fig. 532*a*. The focusing action of this coil, which carries a direct current, is as follows. The axis of the electromagnetic field due to the coil will lie along the axis of the tube, and consequently electrons moving along this axis will experience no force due to the (parallel) magnetic field. Suppose, however, that an electron leaves the axis and acquires a velocity v making an angle α with the axis of the tube, and therefore with the magnetic field (Fig. 532*b*). This electron will have a velocity component $v \cos \alpha$ parallel to the field and $v \sin \alpha$ at right angles. The component parallel to the field can be neglected for the moment, since this gives rise to no force on the electron; but the velocity at right angles to the field will result in a deflecting force. The direction of this force may be seen from Fig. 532*c*. For suppose the direction of the field is down into the paper, the axis of the field passing through the point O, then, by Fleming's left-hand rule, and remembering that a moving electron is equivalent to a current in

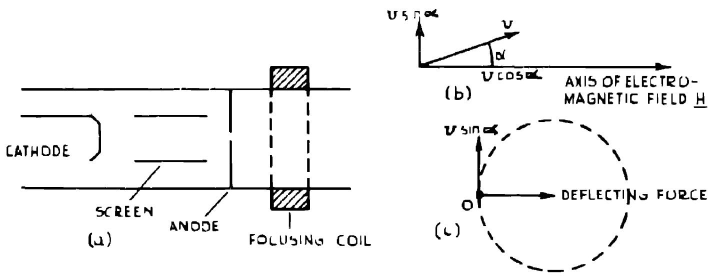


FIG. 532.—Position of focusing coil and principles of electromagnetic focusing.

the reverse direction, it is seen that a force acts on the electron at right angles to its transverse velocity. This force will be constant and equal to $Hev \sin \alpha$ dynes, where e is the charge on the electron in electrostatic units. Under these circumstances, the electron will move in a circle, so far as its transverse motion is concerned. If the radius of its transverse orbit is r cm, and the mass of the electron is m grams, then the centrifugal force will be given by $\frac{m(v \sin \alpha)^2}{r}$ dynes, and this force must be balanced by the deflecting force due to the magnetic field. Thus:—

$$Hev \sin \alpha = \frac{m(v \sin \alpha)^2}{r}$$

i.e.

$$r = \frac{mv \sin \alpha}{He} \text{ centimetres.}$$

The electron is constrained to move in a circle of this radius; its tangential velocity will remain at $v \sin \alpha$ since no force acts on the electron in the direction of this velocity at any time, and consequently there can be no acceleration or deceleration in this

direction. The electron describes one complete revolution of its circular orbit in a time :—

$$t = \frac{2\pi r}{v \sin \alpha} = \frac{2\pi m}{He} \text{ secs.}$$

The time taken for the electron to describe a complete revolution and arrive back on the axis of the magnetic field is independent of both the velocity of the electron and on its departure angle α. During this interval of time, however, the electron will have travelled a distance $\frac{2\pi m}{He} . V$ cm, where V is the axial velocity of the electrons in the beam, and will be considered to have the same value for all the electrons. The motion of the electrons therefore is a spiral one, and if the cathode is considered as a point source of electron emission, the electron stream will be focused to a point at positions distributed at distances $\frac{2\pi mV}{He}$ cm along the axis of the tube. If the distance between cathode and screen is l cm, then a field of strength

$$H = N . \frac{2\pi mV}{el} \text{ gauss}$$

will focus the electron beam accurately on to the screen, where N is a whole number. In practice the smallest field capable of giving a focusing action is employed, so that $N = 1$, and the electrons are brought together on the screen after performing only one revolution in the spiral path.

DEFLECTION

So far, only the general construction of the cathode ray tube has been considered, together with the arrangements for producing on the fluorescent screen a small sharply defined spot of light. It is now necessary to investigate the means by which this spot may be moved across the screen under the influence of externally applied forces. There are two methods, known respectively as electromagnetic and electrostatic deflection. These will now be considered separately.

Electromagnetic deflection

Fig. 533 shows how a pair of deflecting coils may be arranged at the neck of the tube to produce a magnetic field at right angles to the electron beam. The direction of the deflection is given by Fleming's left-hand rule the electron beam being subject to a deflecting force in just the same way as a current-carrying conductor. It is important that the magnetic field produced by the coils be uniform and symmetrical. The deflection is proportional to the magnetic field ; that is, proportional to the current passing through the deflecting coil. Such a coil has an inductive impedance and may therefore disturb the circuit under test.

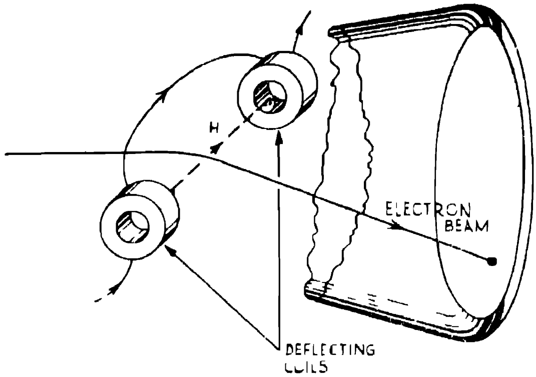


FIG. 533.—Electromagnetic deflection.

Electrostatic deflection

In the electrostatic method of deflection, two plates are arranged one on each side of the beam, as shown in Fig. 534. If a voltage is applied across these deflector plates, the beam will be attracted towards the positive plate and repelled from the negative one, so that the spot of light on the screen will change its position.

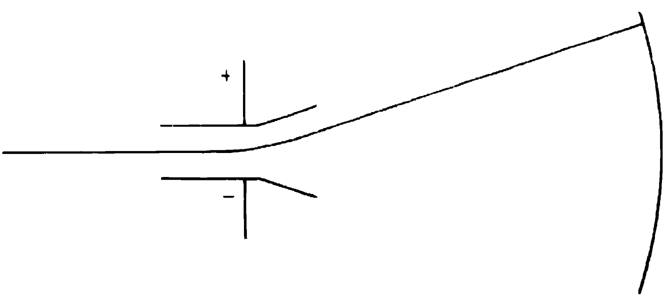


FIG 534.—Electrostatic deflection.

The deflection corresponding to a particular value of deflecting voltage (or current) will be directly proportional to the strength of the deflecting electrostatic (or magnetic) field, to the length of the electron path lying in the field, and to the distance of the fluorescent screen from the deflecting system. For high sensitivity all these constant factors are made as large as possible subject to limitations of space. The deflection also depends on the anode voltage, to which it is inversely proportional, since a higher anode voltage gives an increased electron velocity and hence a smaller deflection. Thus " hard " tubes, which have a higher anode potential than " soft " tubes, are less sensitive. Again, a cold-cathode or high-voltage tube is less sensitive than a hot-cathode or low-voltage tube. In Fig. 534 the deflector plates are very close together at

the end nearer the cathode to give maximum sensitivity, but diverge towards the fluorescent screen in order that a wide angular deflection of the beam may be possible.

If the voltage across the deflector plates (or the current through the deflecting coils) is alternating, the spot will follow the variations in voltage (or current) exactly and without appreciable time lag. As the spot of light is moved back and forth under the influence of an alternating potential applied to the deflector plates, it will trace out a straight line; this, due partly to the normal persistence of vision and partly to the "after-glow" properties of the fluorescent material, will appear as a continuous line, unless the frequency is very low, in which case the actual motion of the spot may be followed by the eye.

TIME BASES

If a second pair of deflector plates are so fitted as to produce an independent deflection at right angles to the first, then the spot of light may be moved to any position on the screen instead of merely in a straight line. The plates that cause a horizontal deflection are called the "X" plates, and those causing a vertical deflection are called the "Y" plates. If the voltage under examination is applied across the Y plates, and at the same time a voltage is applied to the X plates that will cause the spot to travel at a uniform rate across the screen, then clearly the trace on the screen will be an accurate graph showing the instantaneous voltage plotted against time.

If the voltage applied to the Y plates is a transient—that is, if it occurs once only—it may be impossible to examine the trace by eye, and one must therefore photograph the trace as it is produced, or else use a fluorescent screen that will retain the trace for a reasonable period after its formation. If, however, the voltage under examination is recurrent, the voltage on the X plates may be so controlled that it moves the light spot uniformly across the screen from left to right, and having reached the limit of its sweep to the right it then returns the spot very rapidly to the left where it begins its sweep again. If it can be arranged that the second and subsequent traces lie exactly on top of the first, then the eye will obtain the impression of a stationary trace on the screen; Fig. 535 shows such a trace. It will be seen from this example that, during the time that the voltage on the Y plates passes through two complete cycles, the deflecting voltage on the X plates moves the spot once across the screen from left to right, giving the trace *ABCDEF*. Then the spot is moved very rapidly from right to left giving the trace *FA*, which is called the fly-back. This fly-back appears inevitably since the return of the spot, in practice, occupies a finite time; it is, however, much fainter than the main trace, because the fly-back time is small.

The means whereby the deflecting voltage on the X plates is controlled is called a "time base", and in order to ensure that

a stationary trace appears on the screen it is necessary to "synchronise" the time base with the voltage under examination, that is to say, the time occupied by the sweep and the fly-back must be equal to the time occupied by a whole number of cycles of the recurrent voltage on the Y plates. Methods of producing a time base of the required shape, and then synchronising it as required, will be considered separately.

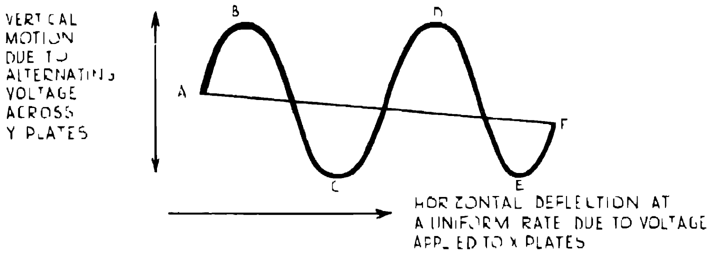


FIG. 535.—Formation of an oscilloscope trace.

The time base that would be required to produce the trace of Fig. 535 is what is known as a "linear" time base; that is, the voltage applied to the X plates takes the form shown in Fig. 536.

Fig. 537 shows a simple time base using a gas relay. This consists of a triode that has been made "soft" by the inclusion, after evacuation of the air has been completed, of a small quantity of an inert gas such as neon, argon or helium; certain devices of this type have been given the trade name of "thyratron". For a given anode voltage and a sufficient negative bias on the grid, this valve will behave as an ordinary triode biased beyond cut-off;

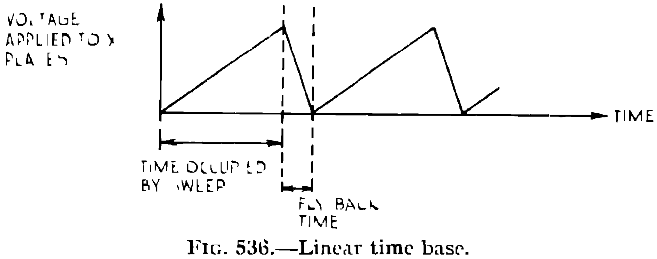


FIG. 536.—Linear time base.

that is, no anode current will flow. Now suppose that the grid bias remains constant whilst the anode voltage is increased; a value of anode voltage will be reached, where anode current begins to flow, and it is here that the similarity to the ordinary "hard" triode ends. For the inert gas, being at low pressure, is readily ionised by the impact of the electron stream on its molecules. At each collision more electrons are released to join the electron stream, and these in turn collide with other molecules and cause further ionisation. In this way a very rapid cumulative current is obtained, giving

a very low resistance between anode and cathode. Once ionisation has taken place and current has started to flow, this condition will persist until the anode voltage is removed—or at any rate reduced to a very low value, much lower than the value of anode voltage at which anode current would cease in a similarly biased "hard" triode.

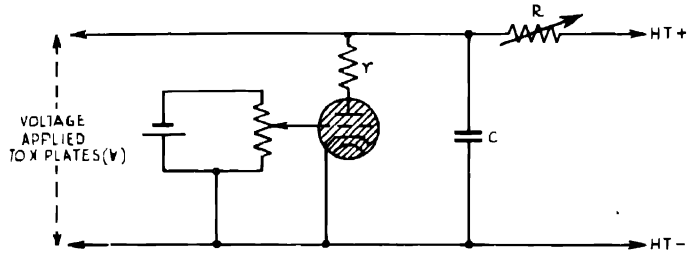


FIG. 537.—Simple time base using a gas relay.

In the circuit of Fig. 537, suppose that the HT voltage is 200 volts and that it has just been switched on. The condenser C will begin to charge up through the resistance R and the voltage V across the condenser will rise exponentially at a rate controlled by the time constant CR. When V reaches a certain value, say 150 volts, the gas relay will "trigger", this triggering point being determined by the negative potential on the grid. Once the relay has triggered, its resistance becomes negligible; and the condenser C discharges through the resistance r, which is a small resistance (about 500 ohms) to limit the anode current in the relay. The value of r is a compromise, since it must be large enough to limit the anode current to a safe value, and at the same time it must be small enough to ensure a rapid fly-back. The voltage V therefore decreases exponentially at a rate determined by the much smaller time constant Cr until it reaches a value of about 10 volts, when the ions in the tube re-associate, and the anode current is once more

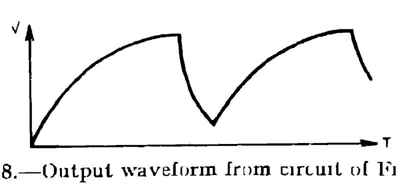


FIG. 538.—Output waveform from circuit of Fig. 537.

cut off by the grid bias. This process repeats itself, the voltage V varying with time, as shown in Fig. 538. It will be noticed that such a time base does not give a strictly linear deflection, but the discrepancy can be minimised by ensuring that the condenser only charges up to about 20 per cent. of the available HT voltage, since the growth of charge on the condenser over this interval is approximately linear. This method of securing a linear time base has the disadvantage of requiring a high value of HT supply voltage. For

suppose that the X plates require a deflecting voltage of 100 volt to give a full-screen deflection, the HT supply required must be, at a minimum, 500 volts.

An alternative method of obtaining a linear scan is shown in Fig. 539. Here the condenser is charged through a saturated diode, which acts as a constant-current device, and therefore the condenser C will charge linearly. If the diode, which must be

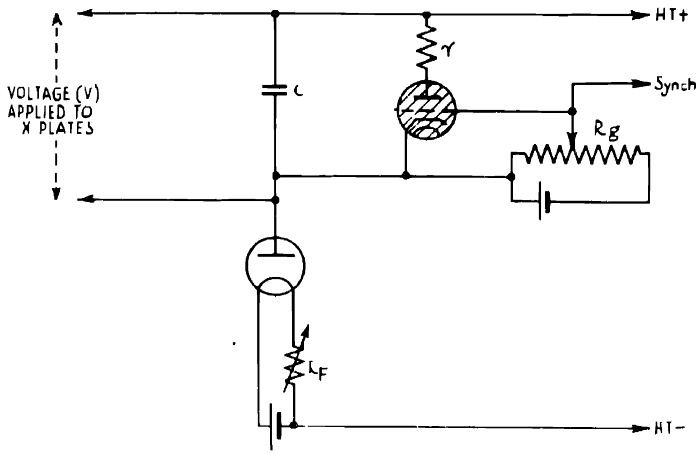


FIG. 539.—Time base circuit using a saturated diode.

directly heated, is given a reduced filament current (controlled by R_F), it is possible to arrange that the emission is so much reduced that the valve saturates for an anode voltage of, say, 20 volts. If, therefore, the anode voltage is raised above 20 volts, however much it may be increased, the anode current will remain constant; and the condenser, charging through the diode, will therefore charge linearly until it attains a voltage within 20 volts of the available HT supply voltage. Thus a linear sweep is obtained provided only that the available HT voltage is about 20 volts in excess of the deflecting voltage required to give a full-scale deflection. The fly-back is obtained by using a gas relay across the condenser as in the previous circuit. The charging rate is controlled by the resistance R_F and the amplitude of the sweep by R_g.

Fig. 540 shows a circuit that operates in much the same way, but here a pentode is used as a constant-current charging device. This method has the advantage that a directly heated valve is not essential, and consequently a separate source of DC LT voltage is not required. The principle of this circuit is that a pentode, having suitable constant voltages applied to screen and grid, has an anode current that is virtually independent of anode voltage provided the latter exceeds, say, 60 V. Thus the voltage on the condenser C increases linearly until it attains a value about 60 V below the available HT supply. The charging rate is controlled by the

potentiometer P, which varies the value of anode current taken by the pentode; and the amplitude of the sweep is controlled by Q, which varies the triggering point of the gas relay.

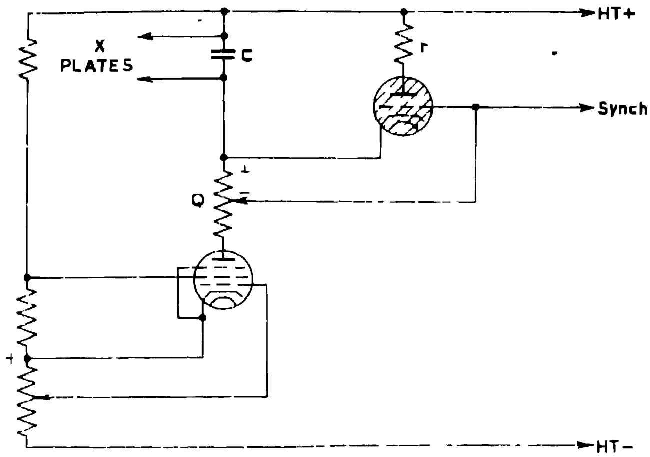


FIG. 540.—Time base circuit using a pentode charging arrangement.

Time base circuits employing gas relays have one common disadvantage, namely, that they cannot be used to produce very high velocity sweeps and so cannot be used with CROs intended for examining high frequency waves. This is due to the comparative slowness with which the ionised gas re-associates after the relay has triggered, since the condenser cannot begin to charge again until re-association is complete. It must be remembered that a saw-tooth waveform contains not only the fundamental, but also an infinite series of harmonics, of which at least the first ten are of sufficient amplitude to be important and require production by the saw-tooth generator. For these reasons, time base circuits of the type described will not operate satisfactorily at frequencies greatly in excess of 10 kc/s.

Time base using a transitron saw-tooth oscillator

Another time base suitable for low and medium frequency work is shown in Fig. 541. This transitron saw-tooth oscillator has the advantage that it uses only one valve.

Consider that at a given instant the valve is drawing a high anode current that is greater than the charging current of C_1 through R_2. The condenser C_1 will discharge, causing a decrease in anode voltage. As the anode voltage falls, the anode current will at first remain practically constant, and then commence to fall as the "knee" of the anode characteristic is passed; this results in a rising screen current which causes a falling screen voltage due to the drop across R_3, and therefore a falling suppressor voltage due to the coupling condenser C_2. These factors cause a further decrease in anode current and the whole process continues

cumulatively until the anode current is completely cut off tl suppressor by this time being at a negative potential

This condition persists for a short time while C_1 begins t charge up exponentially via R_1 and R_2 This means that the potenti of the anode rises and a point is eventually reached where the valv takes anode current again at the expense of the screen causing rise in screen volts and therefore in suppressor volts This proces also is cumulative giving a rapid increase in anode current which

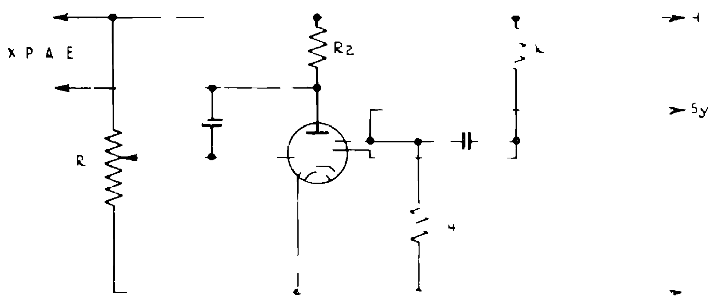


Fig 541 Transitron ... the ... oscillator

discharges C_1 and leaves the suppressor positive If the suppressor grid take a large current when it is driven positive and if C_2 is sufficiently small compared with C_1 the suppressor grid will b driven negative by the voltage drop across R_4 due to suppressor current flowing before the discharge of C_1 is complete This cut off the anode current and C_1 begins to recharge

With suitable choice of components the voltage across C_1 has a saw toothed waveform and may be used to provide a linear time base (see Fig 542)

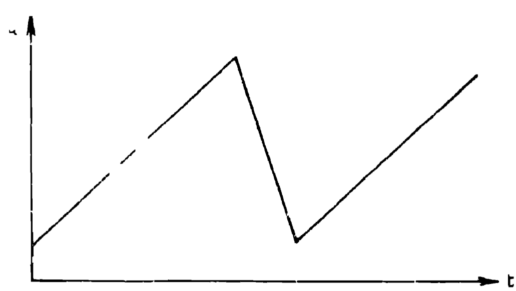


Fig 542 Saw tooth output of transitron oscillator

The potentiometer R_1 controls the total space current and therefore the duration of the periods of charge and discharge, it thus determines the sweep frequency of the time base

Synchronisation of time bases

It has been mentioned that, in order to obtain a stationary trace, the time taken by the time base sweep plus the fly-back time must be equal to the time occupied by a complete number of cycles of the voltage under examination. The frequency of the time base is controllable over wide limits by a coarse control that switches in suitable values of capacity, and by a fine control in the form of the potentiometer that controls the charging current, *e.g.*, P in Fig. 540. The mere adjustment of time base frequency is not, however, sufficient to ensure a stationary picture, for this would demand that the frequency of the time base and of the voltage under examination both remain absolutely constant. It is therefore necessary to employ some device to synchronise the time base with the voltage under examination.

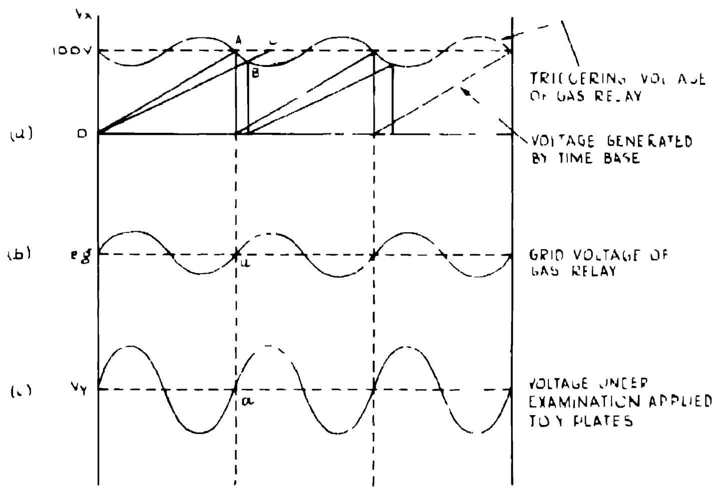


FIG. 543.—Illustrating principle of synchronisation.

If a small portion of the voltage applied to the Y plates is also fed to the grid of the gas relay, it will cause the grid voltage to vary a little about its normal value at the frequency of the voltage under examination. Now suppose that the gas relay is adjusted to trigger at, say, 100 volts, and that the time base frequency has been adjusted as nearly as possible so that the relay triggers at the end of each complete cycle of the voltage on the Y plates (*see* Fig. 543); this should lead to a stationary trace showing one complete cycle (assuming, for simplicity, that the fly-back time is zero).

When the time base is correctly adjusted, the gas relay is triggering at A corresponding to the point α on the Y plate voltage, and to point a on the grid voltage of the gas relay; *i.e.*, the gas relay is triggering at the correct value of 100 V, the grid voltage having been adjusted to give this condition. Now suppose that the

time base frequency falls below its correct value. This means that at the point α on the Y plate voltage curve, the gas relay has not yet triggered because the triggering voltage of the relay (100 V, say) has not yet been attained. Since the grid voltage of the gas relay is varying, however, so also will the triggering voltage of the relay, with the result that the relay will trigger at the point B; that is, at a lower voltage, and consequently earlier, than would have been the case (point C) without the application of a synchronising voltage. A similar argument holds if the time base frequency rises above its correct value, the synchronising voltage this time delaying the triggering of the relay. Provided, therefore, that the time base frequency is reasonably close to the value required, it is possible to arrange that for a very small value of synchronising voltage the relay always triggers within a very small fraction of a cycle of its correct value. Theoretically, however large the discrepancy a sufficiently large synchronising voltage will synchronise the time base perfectly; but large synchronising voltages, though producing a stationary trace, may give rise to a distorted picture. It is therefore important that the synchronising voltage be as small as possible and the time base frequency adjusted with reasonable accuracy, in which case a perfectly stable trace will result.

High-frequency time bases

It has been mentioned that the gas relay or soft valve type of time base does not give satisfactory operation at high frequencies and so for a time base frequency in excess of 10 kc/s it is necessary

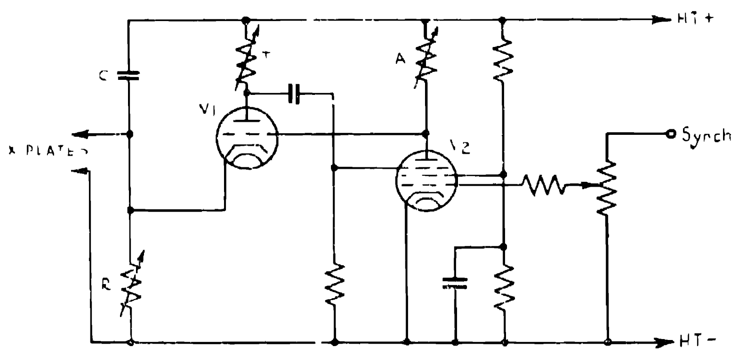


FIG. 544.—Puckle's hard valve time base.

to use a form of time base employing hard valves only. One such circuit, due to Puckle, is shown in Fig. 544, and its operation is somewhat similar to that of the multivibrator relaxation oscillator (*see* p. 490).

To explain the action of this circuit, consider the condenser C uncharged; V_1 will then take no anode current, the anode and cathode both being at the same potential, namely, that of HT+.

The grid of V_1, moreover, will be at a lower potential than the cathode, due to the voltage drop across A caused by the anode current of V_2. In this condition the condenser C will begin to charge exponentially through R; and as the voltage across C increases, the voltage of the cathode of V_1 will fall relative to HT+. Initially the grid of V_1 is at a lower potential than the cathode; and if this bias is made fairly large, by adjustment of A, then the valve V_1 will be cut off until the voltage across C has reached a predetermined value. When V_1 starts to draw anode current, its anode potential drops, and this drop is applied to the suppressor grid of V_2; this causes a rise in anode potential of V_2 which is applied to the grid of V_1, causing a further increase in anode current. Thus the condenser begins to discharge, and the discharge current increases very rapidly, so giving the fly-back. As soon as the condenser has discharged the anode current ceases, and the consequent rise in V_1 anode potential drives the suppressor grid of V_2 positive, and hence the grid of V_1 negative, thus giving a rapid restoration to the original conditions.

This circuit will produce time base frequencies up to about 100 kc/s. The resistance R determines the charging rate of C, and is designated the "velocity" control; it is the fine adjustment on the sweep frequency. Resistance T limits the discharge current and is a control on the fly-back time; it is usually called the "trigger" control. Resistance A determines the negative bias on the grid of V_1 when the sweep is about to start, and so therefore controls the voltage that is being applied to the X plates when V_1 begins to draw anode current. It therefore determines the length of the sweep, and is called the "amplitude" control

Finally, the synchronising voltage is applied to the grid of V_2; suppose that, at a given instant, this is negative. The resultant reduction in anode current of V_2 causes the anode potential of V_2 to rise; this drives the grid of V_1 more positive, thereby reducing the anode potential at which V_1 commences to draw anode current. The explanation of synchronisation as applied to soft valve time bases will therefore apply equally in this case; since, if the time base is running slow, a negative synchronising voltage is being applied to the grid of V_2, and therefore a much larger positive voltage to the grid of V_1, at the instant when V_1 should have triggered. The result is a lowering of the triggering voltage of V_1, so that the triggering takes place almost instantaneously.

ADDITIONAL FEATURES

Deflection amplifiers

If the voltage under examination is applied direct to the Y plates, the voltage range is limited by the sensitivity of the tube. For instance, if the voltage is small, say of value 2 volts RMS, the trace may have an amplitude of only 2 mm. In order to make possible

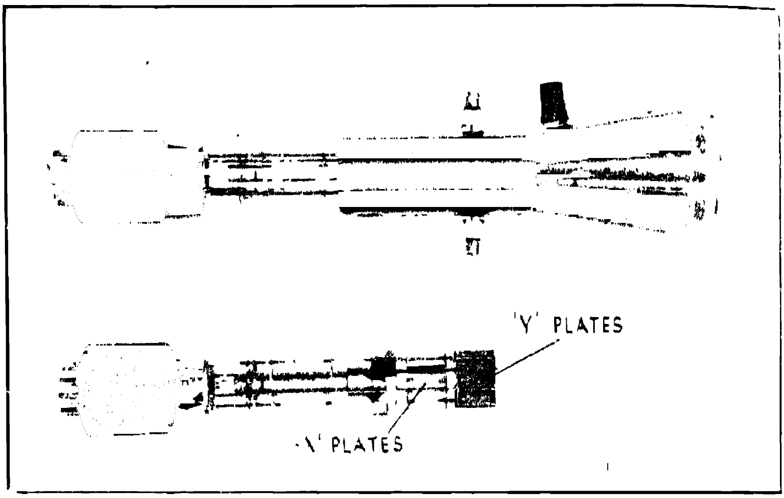


PLATE 29.—Single-beam cathode ray tube.

the examination of small voltages, deflection amplifiers are usually built into the CRO. These are usually straightforward single-stage amplifiers of variable gain, and they enable voltages of very small amplitude to be examined. It is important that a deflection amplifier should introduce as little distortion as possible, since otherwise the CRO becomes valueless as a means of estimating the amount of distortion in a given waveform. It should also be realised that however good a deflecting amplifier may be, it will inevitably introduce distortion at very low and at very high frequencies, so that in general the use of deflection amplifiers limits the frequency range of the instrument.

Attenuators may also be provided for the examination of high voltages in excess of the voltage required for full-screen deflection.

Double-beam tube

A cathode ray tube of the type so far discussed is able to produce only a single trace, showing how a single voltage varies with time; whereas it is often required to compare two such traces. This may easily be done by the use of the double-beam tube due to Fleming-Williams.

Fig. 545 shows how the two beams are obtained. The electrode system is exactly the same as that in a single-beam tube, as are the methods of focusing and brilliancy control. The only difference is that the single beam, after formation, is split by the presence of an earthed

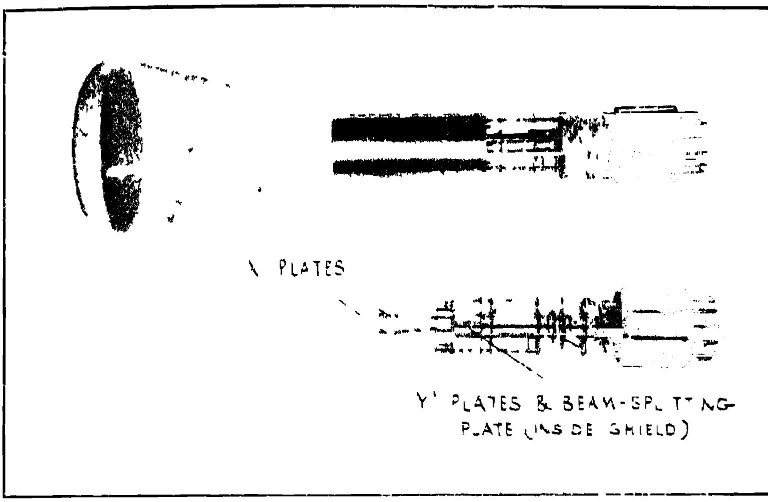


Plate 30.—Double-beam cathode ray tube.

"beam-splitting plate" E mounted between the Y deflector plates. Besides splitting the single beam into two separate beams, this plate acts as a deflecting plate common to the two beams. Thus the first beam may be deflected by applying a voltage between E

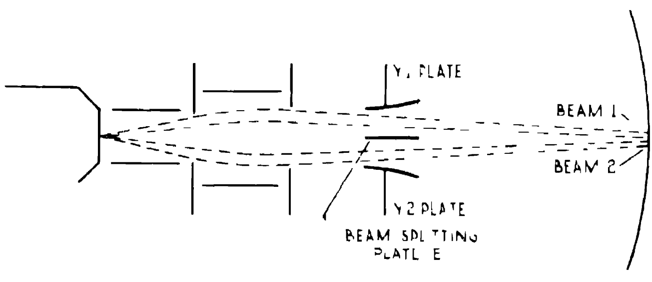


Fig. 545.—Fleming-William's double-beam tube, showing beam-splitting plate.

and Y_1, and the second beam may be deflected independently by a voltage between E and Y_2. The two beams, after formation, pass between normal X plates, which control the sweep of both beams simultaneously by means of a common time base.

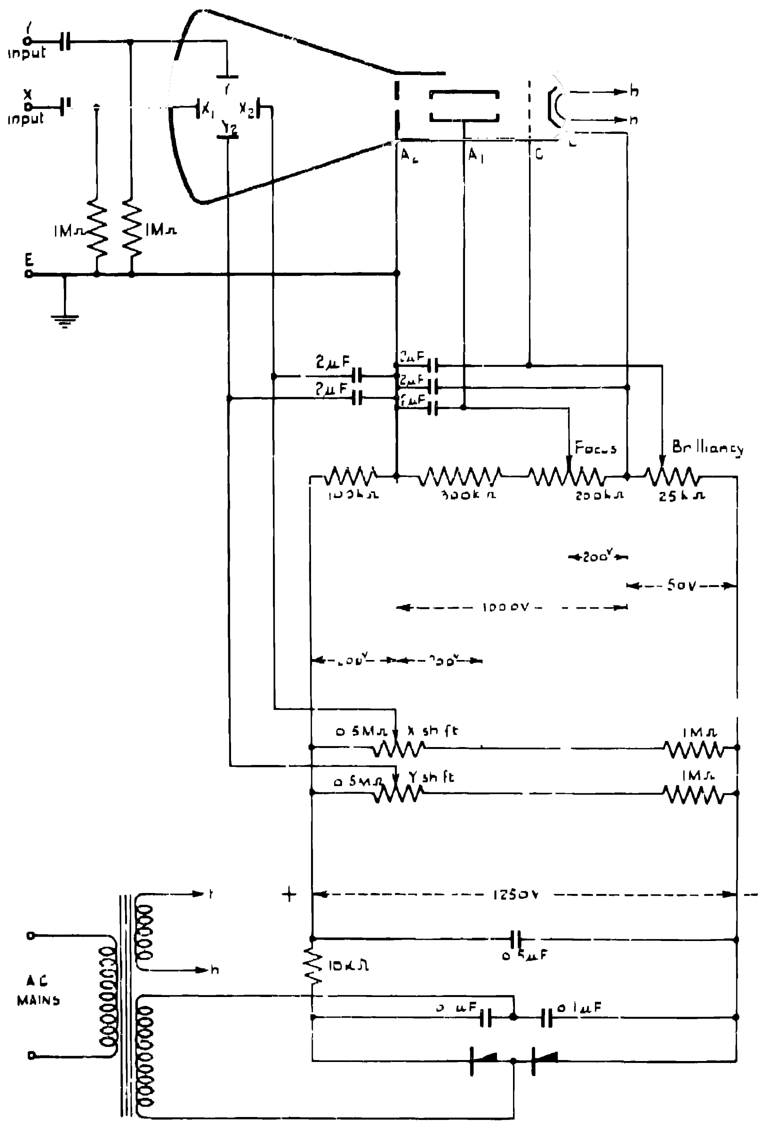


FIG. 546.—Typical power supply for single-beam electrostatically-focused tube.

Shifts

In the double-beam tube it is essential to provide some means of separating the two traces produced on the screen; for, as shown in Fig. 545, the beams when not deflected will be very close together and will give a single spot, or two spots very close together. This separation of the traces is easily accomplished by applying a steady positive potential to Y_1 and Y_2 relative to E. If these potentials are made independently variable, it is possible to move the two traces vertically to any part of the screen. This vertical or Y-shift facility is usually provided also on single-beam tubes. An X-shift is also provided on tubes of both types in the form of a variable steady potential applied to the X plates in addition to the time base voltage. By the use of these two shifts, the trace may be moved to any part of the screen.

Distortion

If the screen is flat, equal increments in deflecting voltage will produce equal deflections. As, in practice, the screen is curved, a small amount of distortion is introduced. With normal tubes, however, this is of the order of 2 per cent., and can be neglected.

Power supply

Fig. 546 shows a typical power supply arrangement for a single-beam electrostatically-focused tube.

APPLICATIONS OF THE CRO

Voltage measurements

The examination of voltage waveforms applied to the Y plates has already been considered in order to describe the working of the CRO. In addition to the examination of the waveform, it is possible to use the CRO to measure the amplitude of the voltage. In order to do this the tube must be calibrated; that is to say, if it is known that a certain voltage produces a definite amplitude of trace, then twice that voltage will produce a trace having twice that amplitude. A CRO may be thus calibrated either for peak values or for RMS values; and if, in addition, the deflection amplifiers and/or attenuators are calibrated, a wide range of fairly accurate measurement can be obtained.

Current measurements

Current waveforms may be examined by passing the current through a non-inductive resistor and examining the voltage drop across this in the normal way. If it is required to *measure* current, the resistance must be small in order to minimise its effect on the circuit under test; it will therefore usually be necessary to use a calibrated amplifier for such a measurement. One particularly useful application of the double-beam tube is that it enables current and voltage tests to be made simultaneously on a circuit, so that

the phase relationship between current and voltage may be seen clearly by comparing the two traces.

Some CROs have electromagnetic deflector coils fitted as an alternative to electrostatic deflector plates; in such a case current measurements can be made directly; but in general the impedance of the deflector coils is such that this method introduces greater errors than does the presence of a small resistance.

Frequency comparison

Frequencies may readily be measured if the frequency of the time base is known. Thus if the time base is sweeping 1000 times per second, and a stationary trace of one cycle appears on the screen, the voltage on the Y plates must have a frequency of 1000 c/s. Similarly, if two full cycles appear on the screen, the frequency must be 2000 c/s. This method will be only an approximate one if the applied signal is allowed to synchronise the time base. Suppose that the frequency were not 2000 c/s, but, say, 2100 c/s; then, if the time base were set to 1000 c/s, it would be synchronised by the work voltage and would actually run at 1050 c/s, giving two complete cycles on a stationary trace.

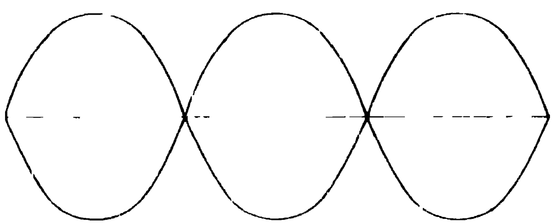


FIG. 547.- Trace when frequency applied to Y plates is 1½ times the sweep frequency.

It is however not essential to use the applied voltage for synchronising the time base. Suppose a 50 c/s mains supply, at a suitably reduced voltage, is applied to the grid of the gas relay, and the time base is set to give a 50 c/s sweep. Then, supposing the mains to give an accurate 50 c/s frequency, the time base will be synchronised to sweep at 50 c/s. If, then, the 2100 c/s input voltage is applied to the Y plates, a stationary trace showing 42 complete cycles will result. Similarly, if the time base were set for 100 c/s and synchronised by the 50 c/s supply, the trace would again be stationary and would contain 21 complete cycles. This can be made the basis of an accurate method of frequency comparison. It is, of course, not necessary that the synchronising voltage should be the 50 c/s mains frequency; it might equally well be obtained from a standard oscillator of, say, 1 kc/s. In order that voltages other than the work voltage may be used for synchronising the time base, it is usual to have a "Synch" terminal which may either be strapped to the Y plates or connected to an external synchronising voltage as required.

In order to obtain a stationary trace it is not necessary that the frequency applied to the Y plates should be an exact multiple of the sweep frequency, although this is the only case where an integral number of cycles is obtained. Suppose that the sweep frequency were 1000 c/s and that a frequency of 1500 c/s were applied to the Y plates. In this case the Y plate voltage would complete 3 cycles while the time base swept twice. The resultant trace would be as shown in Fig. 547.

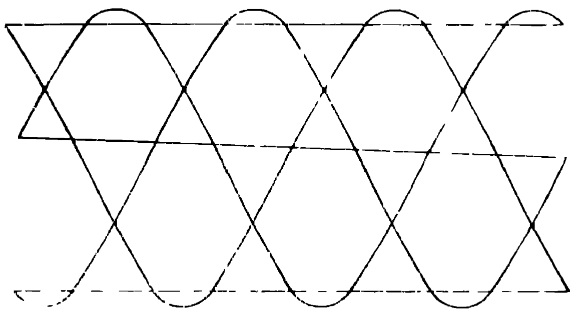


FIG. 548.—Trace when frequency applied to Y plates is $1\frac{1}{2}$ times the sweep frequency.

Similarly, if the ratio of the work frequency to the sweep frequency were 4 : 3, a stationary trace would result; but this time the trace would repeat itself every third sweep of the time base and would be correspondingly fainter and more involved (*see* Fig. 548).

Phase-shift measurements

If the time base circuit is disconnected from the X plates and a sinusoidal voltage $E \sin \omega t$ is applied, the beam will execute

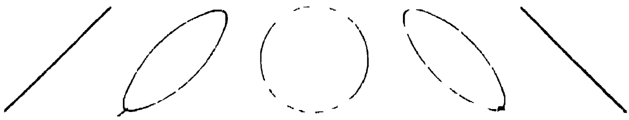


FIG. 549.—Traces illustrating phase measurements.

a horizontal sweep, travelling backwards and forwards in simple harmonic motion. If, simultaneously, a voltage of equal amplitude and frequency but in a different phase, *e.g.* $E \sin (\omega t + \varphi)$, is applied to the Y plates, a stationary trace will result, but the shape of the trace will depend on the phase angle φ between the two voltages.

Fig. 549 shows the shape of this trace for different values of the phase angle φ. Some of these are easily verified; for suppose

$\varphi = 180°$, then the horizontal deflection at a time t is given by :—

$$x = kE \sin \omega t$$

and the vertical deflection (assuming equal sensitivity) by :—

$$y = kE \sin (\omega t + 180°) = - kE \sin \omega t$$

whence the equation of the trace is seen to be the line :—

$$y = - x$$

Again suppose $\varphi = 90°$, in this case :—

$$x = kE \sin \omega t$$

$$y = kE \sin (\omega t + 90°) = kE \cos \omega t$$

whence, eliminating t, the equation is that of the circle :—

$$x^2 + y^2 = k^2E^2$$

For values of phase angle other than 0, $\frac{\pi}{2}$, π, and $\frac{3\pi}{2}$, the equations are again simple to find, but rather more difficult to identify. The resultant trace, however, can in all cases be shown to be an ellipse.

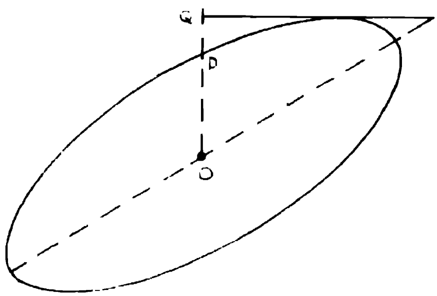


FIG. 550.—Phase ellipse.

From this phase ellipse, the phase angle may be found with fair accuracy. Consider the ellipse of Fig. 550. Let the horizontal and vertical deflections be given by :—

$$x = kE \sin \omega t$$

and $\quad y = kE \sin (\omega t + \varphi)$ respectively.

Then at the point P the horizontal deflection is zero, and therefore $\sin \omega t = 0$.

Hence OP, which represents the vertical deflection at this instant, is given by :—

$$OP = kE \sin \varphi$$

Also OQ is equal to the amplitude of the vertical sweep ;

$$\therefore \quad OQ = kE$$

$$\therefore \quad \sin \varphi = \frac{OP}{OQ}$$

From which φ can at once be determined.

Lissajous figures

Now suppose that the two sinusoidal voltages applied to the X and Y plates are not exactly of the same frequency, but differ by a fraction of a cycle. For instance, suppose that the frequency on the X plate is 1000 c/s, and that on the Y plate is 1000·1 c/s. These frequencies produce a pattern on the screen that approximates to the phase ellipse, but instead of the two voltages having a constant phase difference, this difference is constantly changing, and so therefore will the orientation and eccentricity of the phase ellipse.

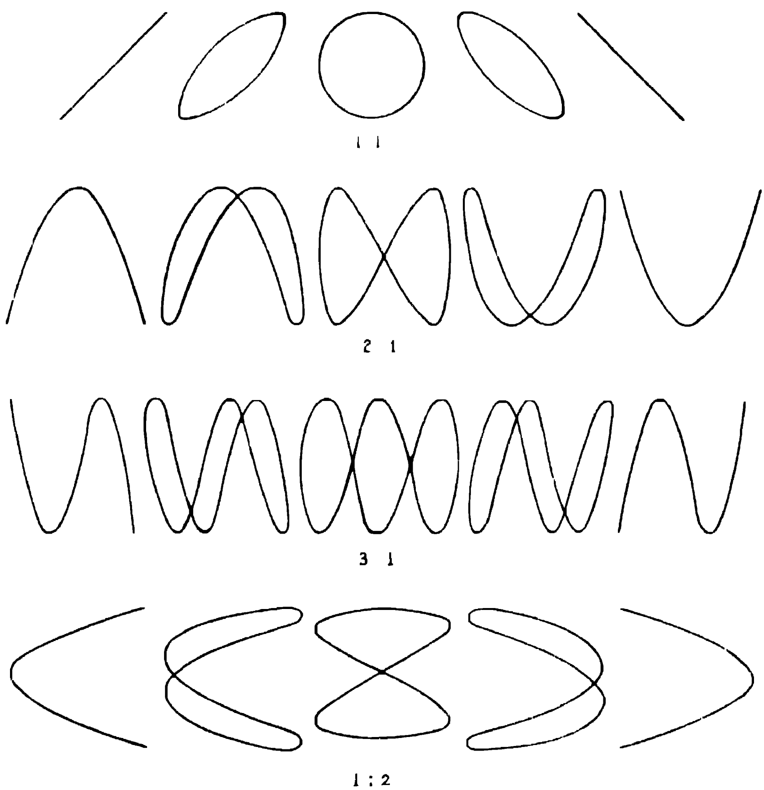


FIG. 551.—Some simple Lissajous figures.

In fact, the trace will pass successively through all the stages shown in Fig. 551 (top line), and in this particular example it will complete the cycle in 10 seconds. In the second line are seen the various possible shapes of the trace for a frequency ratio $f_y : f_x = 2 : 1$. The other traces in this figure should now be self-explanatory. It should be noted that this method for comparing frequencies is impracticable unless the frequencies are an exact multiple to within a few cycles per minute.

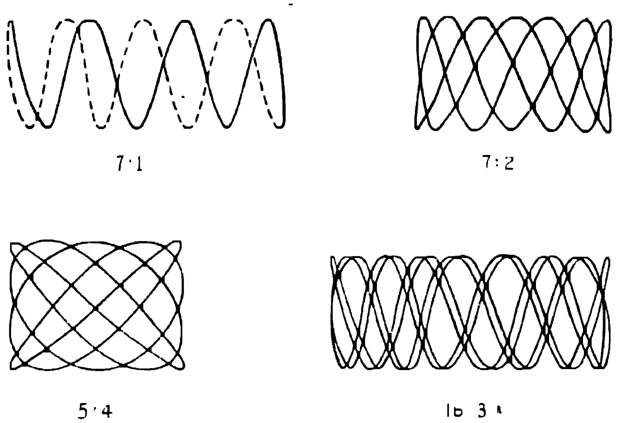


FIG. 552.—Some more complex Lissajous figures.

In order to interpret Lissajous figures it is not necessary to remember pattern shapes, but merely a simple rule connecting the frequency ratio with the number of loops, *viz.* :—

$$\frac{f_y}{f_x} = \frac{\text{Number of loops horizontally}}{\text{Number of loops vertically}}$$

In applying this rule it is simplest to count the loops when the trace is open ; by studying the traces of Figs. 551 and 552, the student should quickly grasp the use of this rule.

Characteristic curves

Another important application of the CRO is in observing characteristic curves. A characteristic curve is a graphical representation of the way in which a given physical quantity y varies with another quantity x. If, therefore a voltage proportional to x be fed to the X plates of a CRO, and at the same time a voltage

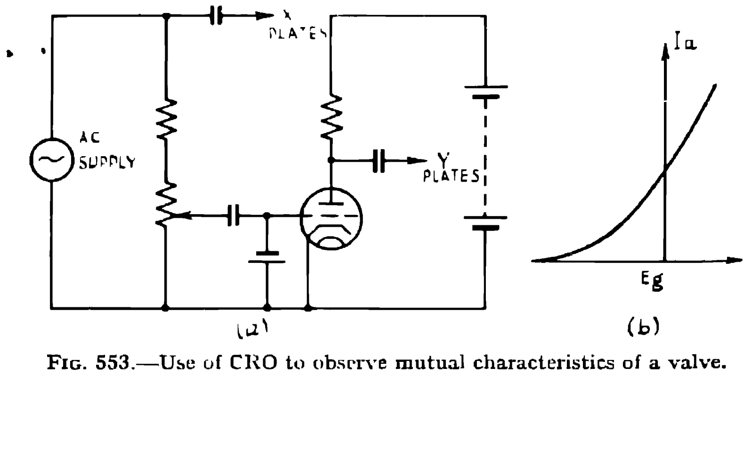


FIG. 553.—Use of CRO to observe mutual characteristics of a valve.

proportional to y is fed to the Y plates, then the spot will trace out the required characteristic.

A particularly easy example of this is a method of obtaining valve characteristics.

Fig. 553 shows a simple circuit enabling the CRO to be used to observe the mutual characteristic of a triode. The valve is set up in the normal way, with suitable DC voltages on the anode and grid, and then an AC voltage of suitable amplitude is applied to the grid; the same voltage is also applied to the X plates of the tube. The anode current will vary with the sinusoidal voltage applied to the grid, and the anode voltage will vary as the anode current. The anode voltage therefore is applied to the Y plates. Provided there is exactly 180° phase-shift in the valve, the trace will show an exact replica of the mutual characteristic of the valve. If the phase-shift has a value differing from 180°, then the spot will travel different paths on its go and return sweeps, and a narrow closed loop will result.

Another characteristic that can be plotted quite simply using a slightly more complex technique is the hysteresis curve of a magnetic material.

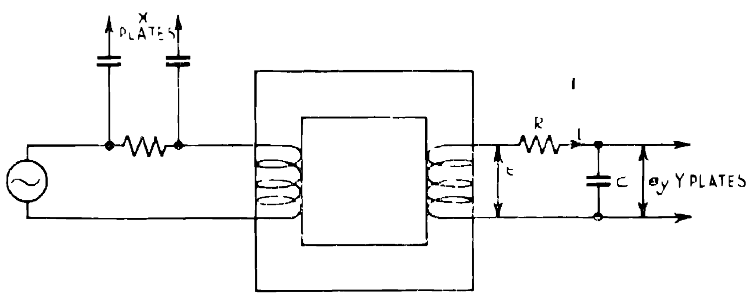


FIG. 554.—Circuit for examining hysteresis curves.

Fig. 554 shows a typical circuit. The specimen of material under examination must effectively form the core of a transformer. The magnetic field H, which is normally plotted on the horizontal axis, is proportional to the magnetising current; an AC supply is accordingly fed to the magnetising windings *via* a small resistance, and the voltage across this is applied to the X plates. The voltage induced across a small secondary winding is proportional to the rate of change of the flux in the magnetic material, that is:—

$$e \propto \frac{dB}{dt}$$

In order to apply a voltage to the Y plates that is proportional to the flux B, it is necessary to use a phase-shifting circuit consisting of capacity C and resistance R arranged as shown.

From the figure, $e_y = \frac{q}{C}$

where q is the charge on the condenser

i.e. $$e_y = \frac{1}{C}\int i dt$$

But if the values of capacity and resistance are so chosen that $\frac{1}{\omega C} \ll R$,

then $$i \doteq \frac{e}{R}$$

$\therefore$ $$e_y \doteq \frac{1}{CR}\int e dt$$

i.e. $$e_y \propto B$$

It should be noted that, in order to produce a hysteresis curve, the amplitude of the magnetising current should be sufficiently large to cause saturation on the peaks, otherwise the trace will merely take the form of a distorted phase ellipse.

CHAPTER 13

FOUR-TERMINAL NETWORKS

A four-terminal network is a network having only one pair of input and one pair of output terminals. When its electrical properties are unaffected by interchanging the input and output terminals, the network is said to be "symmetrical"; if this is not the case, it is said to be "asymmetrical" or "dissymmetrical".

SYMMETRICAL NETWORKS

Symmetrical networks have two important electrical characteristics, namely, characteristic impedance (Z_0) and propagation constant (γ). Two networks having the same characteristic impedance and the same propagation constant are said to be equivalent.

Characteristic impedance

If an infinite number of identical symmetrical networks are connected in tandem as in Fig. 555*a*, the impedance measured at

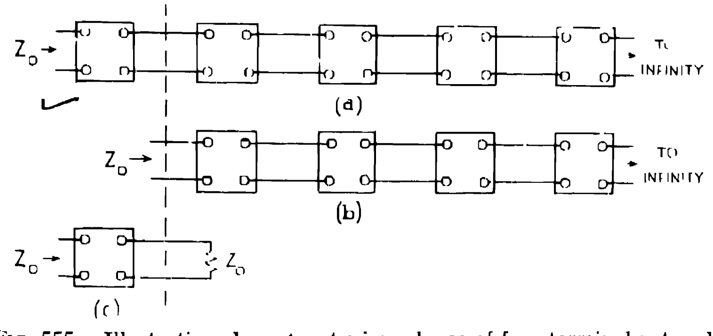


FIG. 555.—Illustrating characteristic impedance of four-terminal network.

the input terminals of the first network will have some definite value depending only on the composition of the networks. This impedance is an important property of the network, and is called its "characteristic impedance", represented by Z_0. It will be seen later that this characteristic impedance may be calculated from a knowledge of the component values of the network.

If the first network of the infinite chain shown in Fig. 555*a* be disconnected, the number of networks remaining will still be

infinite, and therefore the input impedance looking into the second network will be Z_0 (Fig. 555*b*). It follows that, if the first section is connected to an impedance equal to Z_0, as in Fig. 555*c* (instead of to the infinite chain of networks of Fig. 555*b* whose input impedance is Z_0), its input impedance will still be Z_0.

Thus it can be seen that if any symmetrical network is terminated with its characteristic impedance Z_0, the input impedance will also be Z_0. Similarly, if its input terminals are connected to a generator of impedance Z_0, then its output impedance will be equal to Z_0. When both these conditions are satisfied and both the input and output terminals of the network are terminated in Z_0, the network is said to be *correctly terminated*.

Propagation constant

In addition to the characteristic impedance just considered, symmetrical networks have another important property, the propagation constant, which represents the relationship between the input and output currents.

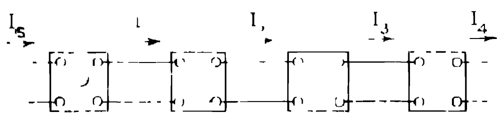


FIG. 556.—Illustrating propagation constant of four-terminal network.

Consider a recurrent network consisting of a series of identical symmetrical sections, as shown in Fig. 556. The current leaving any section will be a definite proportion of that entering the section and will, in general, be out of phase with it. This means that the ratio of the current entering any section to that leaving it is a vector quantity having both modulus and angle. Since all sections are identical, this vector will be the same for all sections. It is convenient to write this vector in the form e^{γ}, where γ is a complex number. Hence let :—

$$\frac{I_S}{I_1} = e^{\gamma} \tag{1}$$

Since each section is identical, it follows that :—

$$e^{\gamma} = \frac{I_S}{I_1} = \frac{I_1}{I_2} = \frac{I_2}{I_3}, \text{ etc.}$$

Thus $$\frac{I_S}{I_2} = \frac{I_1}{I_2} \times \frac{I_S}{I_1} = e^{2\gamma}$$

and $$\frac{I_S}{I_3} = \frac{I_2}{I_3} \times \frac{I_1}{I_2} \times \frac{I_S}{I_1} = e^{3\gamma}$$

Considering the case of a finite number of sections n, correctly terminated, if the current at the sending end is I_S and that at the

receiving end I_R, then :—

$$\frac{I_S}{I_R} = e^{n\gamma} \tag{2}$$

Since γ is a complex number, let $\gamma = \alpha + j\beta$

Then $e^{\gamma} = e^{\alpha + j\beta} = e^{\alpha} \cdot e^{j\beta}$

$$= e^{\alpha} (\cos \beta + j \sin \beta)$$

$$= e^{\alpha} \sqrt{\cos^2 \beta + \sin^2 \beta} \;\Big/ \tan^{-1} . \frac{\sin \beta}{\cos \beta}$$

$$= e^{\alpha} \;\underline{/ \beta} \tag{3}$$

e^{γ} is seen to be a vector of modulus e^{α} and angle β. e^{α} gives the ratio of the absolute magnitudes of currents entering and leaving a section; β gives the phase angle between these two currents.

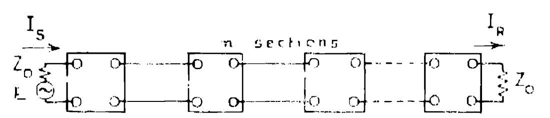


FIG. 557.—n four-terminal networks, correctly terminated.

For n sections correctly terminated :—

$$\frac{I_S}{I_R} = e^{n\gamma} = e^{n\alpha} \;\underline{/ n\beta} \tag{4}$$

In this case the ratio of absolute magnitudes of current sent and current received will be $e^{n\alpha}$, and the phase angle will be $n\beta$. Since the sections are symmetrical, both the input and output circuits will be terminated in the same impedance Z_0. The input voltage E_S will be equal to $I_S Z_0$, and the output voltage E_R will be equal to $I_R Z_0$.

Hence $$\frac{I_S}{I_R} = \frac{I_S Z_0}{I_R Z_0} = \frac{E_S}{E_R}$$

and $$\frac{E_S}{E_R} = e^{n\gamma} = e^{n\alpha} \;\underline{/ n\beta} \tag{5}$$

Attenuation constant and the neper

The real part α of the propagation constant γ is called the "attenuation constant" of the section, and is measured in "nepers".* It is equal to the logarithm, to the base e, of the

* It must be noted that, throughout this and subsequent chapters, the values obtained for attenuation and phase-shift, unless otherwise stated, are in nepers and radians. These results can be converted into the more convenient units for practical work, the decibel and the degree, by multiplying by 8·686 and 57·3 respectively. (*See* also Conversion Tables, pages 839 and 804.)

ratio of the modulus of the current entering the section, to that leaving it.

For $$e^{\alpha} = \left|\frac{I_S}{I_1}\right|$$

$$\therefore \quad \alpha = \log_e \left|\frac{I_S}{I_1}\right| \text{ nepers}$$

For n sections, from equation 4 :—

$$\left|\frac{I_S}{I_R}\right| = e^{n\alpha}$$

so that the attenuation introduced by n sections is :—

$$\log_e \left|\frac{I_S}{I_R}\right| = n\alpha \text{ nepers.}$$

Phase constant

The imaginary part β of the propagation constant γ is called the " phase constant " of the section, and is equal to the angle in radians by which the current leaving the section lags behind that entering it From equation 4 the phase shift introduced by n sections is $n\beta$ radians.

ASYMMETRICAL NETWORKS

Asymmetrical networks generally have different characteristic impedances on the two sides ; and to be strictly accurate, when dealing with such networks, the terms " iterative impedance " and " image impedance " should be used in place of " characteristic impedance ".

Iterative impedance

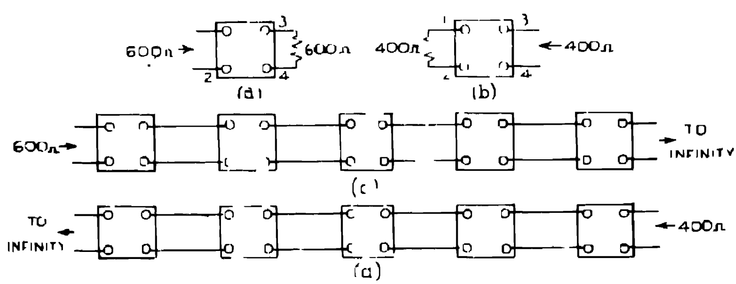


FIG 558 —Illustrating iterative impedances of asymmetrical network

The iterative impedance of a four-terminal network is defined as the input impedance measured at one pair of terminals when an infinite number of such networks are connected in tandem. This is the value of the impedance measured at one pair of terminals

of the network when the other pair of terminals is terminated with an impedance of the same value. Iterative impedances will be different for the two pairs of terminals of an asymmetrical network (*see* Fig. 558). Thus the impedance looking into terminals 1 and 2 is 600 Ω when terminals 3 and 4 are terminated in 600 Ω, and the impedance looking into terminals 3 and 4 is 400 Ω when terminals 1 and 2 are terminated in 400 Ω.

When the two iterative impedances are equal (as they are in the case of symmetrical networks), the common value is the characteristic impedance of the network.

Image impedance

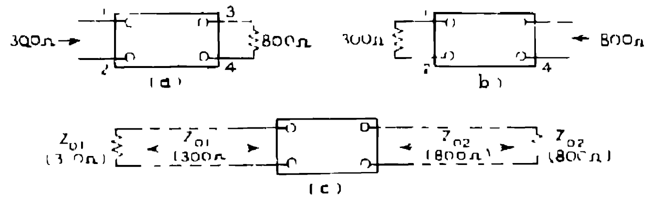


FIG 559 —Illustrating image impedances of asymmetrical network

The image impedances of a network are those impedances such that when one of them is connected across the appropriate pair of terminals of the network, the other is presented by the other pair of terminals (*see* Fig. 559*a* and *b*). In the case of an asymmetrical network, the two image impedances are different. Thus the input impedance at terminals 1 and 2 of the network shown in Fig. 559 is 300 Ω when terminals 3 and 4 are terminated in 800 Ω, and the input impedance at terminals 3 and 4 is 800 Ω when terminals 1 and 2 are terminated in 300 Ω. When the two image impedances are equal (as they are in the case of symmetrical networks), their common value is equal to the characteristic impedance of the network.

An asymmetrical network is said to be correctly terminated when it is terminated in its image impedances. (Fig. 559*c*.)

Image-transfer constant

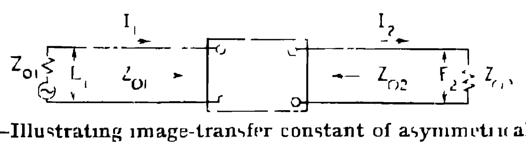


FIG 560 —Illustrating image-transfer constant of asymmetrical network.

When an asymmetrical section, terminated in its image impedances, is considered (Fig. 560), the ratio $\frac{I_1}{I_2}$ will be different

from the ratio $\frac{E_1}{E_2}$. In such a case, the term "propagation constant" is not employed, and instead the "image-transfer constant" θ is considered.

θ is defined as one-half the logarithm to the base e of the vector ratio of the volt-amperes entering the network, to the volt-amperes leaving it, the network being terminated by its image impedances.

Thus
$$e^{\theta} = \sqrt{\frac{E_1 I_1}{E_2 I_2}} \tag{6}$$

The real part of the image-transfer constant is known as the "image-attenuation constant", and the imaginary part is known as the "image-phase constant".

INSERTION LOSS OF A FOUR-TERMINAL NETWORK

When a network is introduced between a generator and a load, the resultant reduction in power in the load is known as the "insertion loss" of the network. It is usual to express this loss of power in either the decibel or the neper notation.

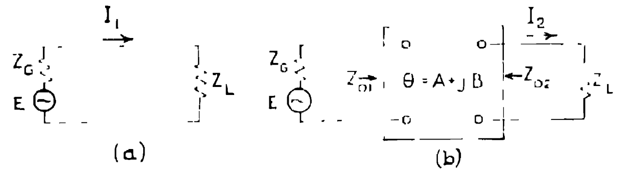


FIG. 561.—Insertion loss of four-terminal network.

Consider the case of a generator of internal impedance Z_g working into a load Z_L (*see* Fig. 561*a*). Let the current flowing be I_1. If a four-terminal network having image impedances Z_{01} and Z_{02} and image-transfer constant θ be inserted between the generator and the load, as in Fig. 561*b*, then the current will be altered to some value I_2, say. The insertion loss is given by :—

$$\text{Insertion loss} = \log_e \left|\frac{I_1}{I_2}\right| \text{ nepers}$$

or :—

$$\text{Insertion loss} = 20 \log_{10} \left|\frac{I_1}{I_2}\right| \text{ decibels.}$$

It can be shown that, if $\theta = A + jB$, the insertion loss is given by :—

$$A + \log_e \left|\frac{Z_g + Z_{01}}{2\sqrt{Z_g Z_{01}}}\right| + \log_e \left|\frac{Z_L + Z_{02}}{2\sqrt{Z_L Z_{02}}}\right| - \log_e \left|\frac{Z_g + Z_L}{2\sqrt{Z_g Z_L}}\right|$$
$$+ \log_e \left|1 - \frac{Z_{01} - Z_g}{Z_{01} + Z_g} \cdot \frac{Z_{02} - Z_L}{Z_{02} + Z_L} \cdot e^{-2\theta}\right| \text{ nepers} \tag{7}$$

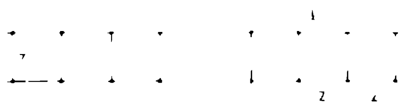
and the insertion phase-shift is given by: —

$$\left\{ B + \text{angle of } \frac{Z_G+Z_{01}}{2\sqrt{Z_G Z_{01}}} + \text{angle of } \frac{Z_L+Z_{02}}{2\sqrt{Z_L Z_{02}}} - \text{angle of } \frac{Z_G+Z_L}{2\sqrt{Z_G Z_L}} + \text{angle of}\left(1 - \frac{Z_{01}-Z_G}{Z_{01}+Z_G}\cdot\frac{Z_{02}-Z_L}{Z_{02}+Z_L}\cdot e^{-2\theta}\right)\right\} \text{ radians} \quad (8)$$

RECURRENT NETWORKS

Ladder networks

The type of recurrent network most commonly encountered in line transmission is the ladder network. It exists in two forms — the "unbalanced" ladder network (see Fig. 562*a*), and the "balanced" ladder network (Fig. 562*b*).

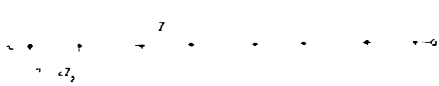


(*a*) Unbalanced (*b*) Balanced

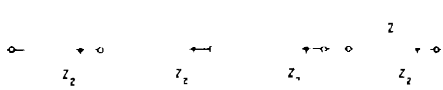
FIG. 562 Ladder networks

Both the balanced and the unbalanced ladder networks may be considered as being built up of a number of "sections" (as shown in Figs. 563 and 564), which are known, by reason of their shape, as "T", "π" and "L" sections in the unbalanced form, and as "H", "O" and "C" sections in the balanced form. These sections are all arranged to have a total series impedance Z_1 and a total shunt impedance Z_2.

(*a*) Ladder network composed of T sections

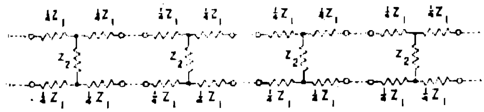


(*b*) Ladder network composed of π sections.

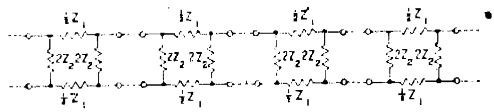


(*c*) Ladder network composed of L sections

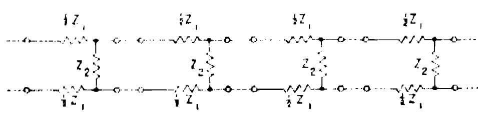
FIG. 563 —Unbalanced ladder networks of Fig. 562*a* represented as a series of sections.



(*a*) Ladder network composed of H or balanced T sections.



(*b*) Ladder network composed of O or balanced π sections.



(*c*) Ladder network composed of C or balanced L sections.

FIG. 564.—Balanced ladder network of Fig. 562*b* represented as a series of sections.

In these figures the T and π sections, of which the ladder network is considered to consist, are shown as being symmetrical; the ladder network could, however, equally well be represented by a series of asymmetrical sections, such as those shown in Fig. 565.

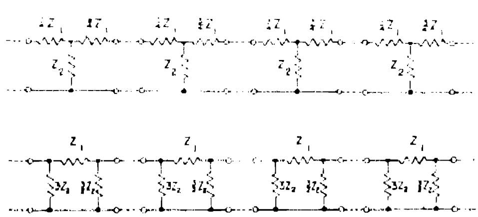


FIG. 565.—Unbalanced ladder network represented as a series of asymmetrical sections.

In fact, the L section (which is asymmetrical) is merely a particular case of the asymmetrical T section (with one series arm equal to zero), or of the asymmetrical π section (with one shunt arm equal to infinity).

Other recurrent networks

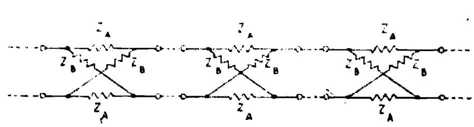
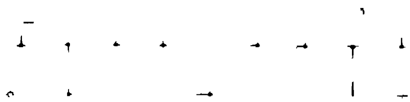
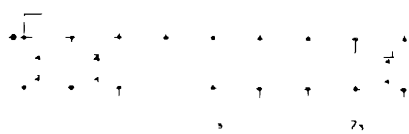


FIG. 566.—Lattice network.

In addition to the ladder structure, two other forms of recurrent networks are encountered, namely the "lattice" and the "bridged-T" networks, shown in Figs 566 and 567 respectively.



(a) Unbalanced.



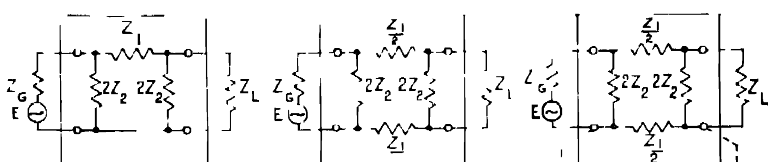
(b) Balanced.

FIG 567.—Unbalanced and balanced forms of bridged-T section.

The lattice section is usually a balanced symmetrical structure. The bridged-T section may be balanced or unbalanced, symmetrical or asymmetrical, though the unbalanced symmetrical form shown is the most usual.

Equivalence of balanced and unbalanced sections

Both the balanced and unbalanced sections have identical transmission properties, as long as no connections are made between the input and output terminals external to the network. Thus in Fig 568, a and b are equivalent, provided that no connection is made as in c.



(a) UNBALANCED (b) BALANCED (c)

FIG 568.—Illustrating equivalence of balanced and unbalanced sections

THE T SECTION

The symmetrical T section shown in Fig 569 is one of the most

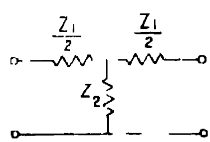


FIG. 569.—Symmetrical T section.

important networks encountered in line transmission theory. It was shown in Fig. 563*a* that a ladder network could be regarded as being made up of these sections. To give a total series-arm impedance of Z_1 in the ladder network, the two series-arm impedances in the T section must each be $\frac{Z_1}{2}$.

Characteristic impedance

To find the characteristic impedance (Z_0) of such a section, terminate the section on one side with Z_0, and determine the input impedance. Equating this input impedance to Z_0 gives an equation from which Z_0 may be determined.

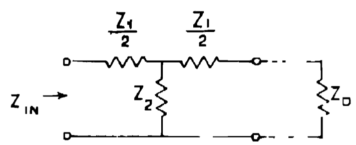


FIG 570.— Symmetrical T section terminated in Z_0

Considering Fig. 570 :—

$$Z_{IN} = \frac{Z_1}{2} + \frac{Z_2\left(\frac{Z_1}{2} + Z_0\right)}{Z_2 + \frac{Z_1}{2} + Z_0}$$

But $$Z_{IN} = Z_0$$

$$\therefore \quad Z_0 = \frac{Z_1}{2} + \frac{Z_2\left(\frac{Z_1}{2} + Z_0\right)}{Z_2 + \frac{Z_1}{2} + Z_0}$$

$$\therefore \quad Z_0Z_2 + \frac{Z_0Z_1}{2} + Z_0^2 = \frac{Z_1Z_2}{2} + \frac{Z_1^2}{4} + \frac{Z_0Z_1}{2} + \frac{Z_1Z_2}{2} + Z_0Z_2$$

$$\therefore \quad Z_0^2 = \frac{Z_1^2}{4} + Z_1Z_2 \tag{9}$$

Thus giving $$Z_0 = \sqrt{\frac{Z_1^2}{4} + Z_1Z_2} \tag{10}$$

Open- and short-circuit impedances

The values of Z_1 and Z_2 may be determined by measuring the input impedance for two given terminations. For convenience, these two terminations are taken as an open-circuit and a short-circuit.

Let the input impedance on open-circuit (Fig. 571*a*) be Z_{OC}.

Then $$Z_{OC} = \frac{Z_1}{2} + Z_2 \tag{11}$$

Let the input impedance on short-circuit (Fig. 571*b*) be Z_{sc}.

Then $$Z_{sc} = \frac{Z_1}{2} + \frac{\dfrac{Z_1 Z_2}{2}}{\dfrac{Z_1}{2} + Z_2}$$

$$\therefore \quad Z_{sc} = \frac{Z_1 Z_2 + \dfrac{Z_1^2}{4}}{\dfrac{Z_1}{2} + Z_2} \qquad (12)$$

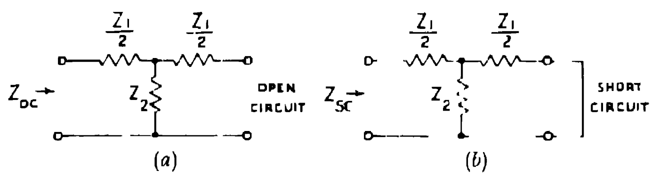


FIG. 571.—Symmetrical T section on open- and short-circuit.

It will be noted that multiplying (11) and (12) gives :—

$$Z_{oc} \,.\, Z_{sc} = \left(\frac{Z_1}{2} + Z_2\right) \times \frac{\dfrac{Z_1^2}{4} + Z_1 Z_2}{\dfrac{Z_1}{2} + Z_2}$$

$$= \frac{Z_1^2}{4} + Z_1 Z_2$$

$$= Z_0^2 \text{ (from equation 10)}$$

that is $$Z_0 = \sqrt{Z_{oc} Z_{sc}} \qquad (13)$$

This formula is most useful, and should be memorised.

From these equations, Z_1 and Z_2 may be determined in terms of Z_{oc} and Z_{sc} :—

Squaring (11) :—

$$Z_{oc}^2 = \frac{Z_1^2}{4} + Z_1 Z_2 + Z_2^2$$

$$= Z_0^2 + Z_2^2 \qquad \text{(from equation 10)}$$

$$= Z_{oc} Z_{sc} + Z_2^2 \qquad \text{(from equation 13)}$$

$$\therefore \quad Z_2^2 = Z_{oc}^2 - Z_{oc} Z_{sc}$$

$$\therefore \quad Z_2 = \sqrt{Z_{oc}(Z_{oc} - Z_{sc})} \qquad (14)$$

$\therefore$ from (11), $$Z_1 = 2\left[Z_{oc} - \sqrt{Z_{oc}(Z_{oc} - Z_{sc})}\right] \qquad (15)$$

Example 1.—

Find the characteristic impedance of the T section shown in Fig. 572

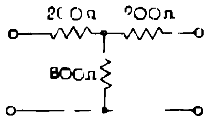


Fig 572

This is most easily done by using the formula $Z_0^2 = Z_{oc} Z_{sc}$

$$Z_{oc} = 200 + 800 = 1000\,\Omega$$
$$Z_{sc} = 200 + (200 \text{ and } 800 \text{ in parallel})$$
$$= 200 + 160 = 360\,\Omega$$

Hence $Z_{oc} Z_{sc} = 1000 \times 360 = 36 \times 10^4$

$$Z_0 = \sqrt{Z_{oc} Z_{sc}}$$
$$= 600\,\Omega \qquad \textit{Ans}$$

Example 2.—

A symmetrical T section composed of pure resistances has the following values for open and short circuit impedances:—

$$Z_{oc} = 800 \text{ ohms } \angle 0^\circ, \qquad Z_{sc} = 600 \text{ ohms } \angle 0^\circ$$

Determine Z_1 and Z_2 for this T section

$$Z_1 = 2\,[800 - \sqrt{800(800-600)}]$$
$$= 2\,[800 - \sqrt{800 \times 200}]$$
$$= 2\,[800 - 400] = 800 \text{ ohms}$$
$$Z_2 = \sqrt{800 \times 200} = 400 \text{ ohms} \qquad \textit{Ans}$$

Propagation constant

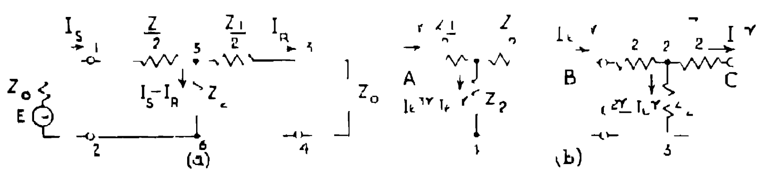


Fig 573.—Propagation constant of T section

Consider a T section correctly terminated, as shown in Fig 573*a*. Let the input current be I_S and the output current be I_R

By definition,

$$\frac{I_S}{I_R} = e^{\gamma}$$

Applying Kirchhoff's Law to mesh 5, 3, 4, 6 —

$$-(I_S - I_R)\,Z_2 + \frac{I_R Z_1}{2} + I_R Z_0 = 0$$

$$\therefore \qquad I_S Z_2 = I_R \left(Z_2 + \frac{Z_1}{2} + Z_0\right)$$

Hence
$$e^{\gamma} = \frac{I_S}{I_R} = \frac{Z_2 + \frac{Z_1}{2} + Z_0}{Z_2}$$

$$\therefore \qquad e^{\gamma} = 1 + \frac{Z_1}{2Z_2} + \frac{Z_0}{Z_2} \tag{16}$$

Since
$$Z_0 = \sqrt{\frac{Z_1^2}{4} + Z_1 Z_2}$$

$$e^{\gamma} = 1 + \frac{Z_1}{2Z_2} + \frac{1}{Z_2}\sqrt{\frac{Z_1^2}{4} + Z_1 Z_2}$$

i.e.
$$e^{\gamma} = 1 + \frac{Z_1}{2Z_2} + \sqrt{\left(\frac{Z_1}{2Z_2}\right)^2 + \frac{Z_1}{Z_2}} \tag{17}$$

The propagation constant of a T section is therefore given by —

$$\gamma = \log_e \left\{1 + \frac{Z_1}{2Z_2} + \sqrt{\left(\frac{Z_1}{2Z_2}\right)^2 + \frac{Z_1}{Z_2}}\right\} \tag{18}$$

From equation 17 —

$$e^{\gamma} = 1 + \frac{Z_1}{2Z_2} + \sqrt{\left(\frac{Z_1}{2Z_2}\right)^2 + \frac{Z_1}{Z_2}}$$

whence
$$e^{-\gamma} = \frac{1}{1 + \frac{Z_1}{2Z_2} + \sqrt{\left(\frac{Z_1}{2Z_2}\right)^2 + \frac{Z_1}{Z_2}}}$$

which simplifies to

$$e^{-\gamma} = 1 + \frac{Z_1}{2Z_2} - \sqrt{\left(\frac{Z_1}{2Z_2}\right)^2 + \frac{Z_1}{Z_2}} \tag{19}$$

Adding

$$e^{\gamma} + e^{-\gamma} = 2 + \frac{Z_1}{Z_2}$$

$$\frac{e^{\gamma} + e^{-\gamma}}{2} = 1 + \frac{Z_1}{2Z_2}$$

or
$$\cosh \gamma = 1 + \frac{Z_1}{2Z_2} \tag{20}$$

Equation 20 may be derived more simply by considering two successive T sections in a recurrent network (*see* Fig 573*b*)

Let the currents at the points A, B and C be —

$$Ie^{3\gamma},\ Ie^{2\gamma} \text{ and } Ie^{\gamma}.$$

Applying Kirchhoff's Law to mesh 1, 2, 3, 4 gives :—

$$(Ie^{2\gamma})\ Z_1 + (Ie^{2\gamma} - Ie^{\gamma})Z_2 - (Ie^{3\gamma} - Ie^{2\gamma})Z_2 = 0$$

Dividing by $Ie^{2\gamma}$ —

$$Z_1 + Z_2 - e^{-\gamma}Z_2 - e^{\gamma}Z_2 + Z_2 = 0$$

$$Z_2(e^{\gamma} + e^{-\gamma}) = Z_1 + 2Z_2$$

$$\therefore \quad \frac{e^{\gamma} + e^{-\gamma}}{2} = 1 + \frac{Z_1}{2Z_2}$$

Thus, as before, $$\cosh\gamma = 1 + \frac{Z_1}{2Z_2}$$

Other useful expressions involving γ may be obtained

Equation 16 gives — $$e^{\gamma} = 1 + \frac{Z_1}{2Z_2} + \frac{Z_0}{Z_2}$$

Equation 20 gives —

$$\cosh\gamma = 1 + \frac{Z_1}{2Z_2}$$

But $e^{\gamma} = \cosh\gamma + \sinh\gamma$

Therefore by subtraction —

$$\sinh\gamma = \frac{Z_0}{Z_2} \quad \text{or} \quad Z_2 = \frac{Z_0}{\sinh\gamma} \qquad (21)$$

Hence $$\tanh\gamma = \frac{\sinh\gamma}{\cosh\gamma} = \frac{Z_0}{Z_2\left(1 + \frac{Z_1}{2Z_2}\right)} = \frac{Z_0}{Z_2 + \frac{Z_1}{2}}$$

Now $$Z_0 = \sqrt{Z_{oc}Z_{sc}}$$

and $$Z_2 + \frac{Z_1}{2} = Z_{oc}$$

$$\tanh\gamma = \frac{\sqrt{Z_{oc}Z_{sc}}}{Z_{oc}} = \sqrt{\frac{Z_{sc}}{Z_{oc}}} \qquad (22)$$

Equation 22 is useful as it enables γ to be calculated from the open and short circuit impedances. It can thus be seen that both the characteristic impedance and the propagation constant of a network can be determined from Z_{oc} and Z_{sc}. From this it follows that two networks will behave similarly if they have the same Z_{oc} and Z_{sc}.

Since $$\cosh\gamma = 1 + \frac{Z_1}{2Z_2}$$

$$\frac{Z_1}{Z_2} = 2(\cosh\gamma - 1) = 2 \times 2\sinh^2\frac{\gamma}{2}$$

$$\sinh\frac{\gamma}{2} = \tfrac{1}{2}\sqrt{\frac{Z_1}{Z_2}} \qquad (23)$$

Now $\sinh\frac{\gamma}{2}\cdot\cosh\frac{\gamma}{2}=\frac{\sinh\gamma}{2}=\frac{Z_0}{2Z_2}$ (from equation 21)

$$\therefore \qquad \cosh\frac{\gamma}{2}=\frac{Z_0}{2Z_2}\times\frac{1}{\sinh\frac{\gamma}{2}}=\frac{Z_0}{2Z_2}\times 2\sqrt{\frac{Z_2}{Z_1}}=\frac{Z_0}{\sqrt{Z_1Z_2}}$$

$$\therefore \qquad \tanh\frac{\gamma}{2}=\frac{\sinh\frac{\gamma}{2}}{\cosh\frac{\gamma}{2}}=\frac{1}{2}\sqrt{\frac{Z_1}{Z_2}}\times\frac{\sqrt{Z_1Z_2}}{Z_0}=\frac{Z_1}{2Z_0}$$

$$\therefore \qquad Z_1=2Z_0\tanh\frac{\gamma}{2} \tag{24}$$

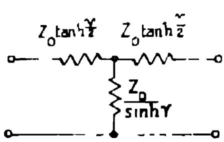


FIG. 574.—T section, showing values of components in terms of characteristic impedance and propagation constant.

Equations 21 and 24 enable the components of a T section to be calculated if Z_0 and γ are known. The section is shown in Fig. 574.

Input impedance of a T section terminated in Z_R

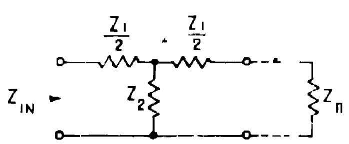


FIG. 575.—T section terminated in an impedance Z_R.

Consider a T section terminated in Z_R (*see* Fig. 575). The input impedance is given by :—

$$=\frac{Z_1}{2}+\frac{Z_2\left(\frac{Z_1}{2}+Z_R\right)}{Z_2+\frac{Z_1}{2}+Z_R}$$

$$\frac{Z_R\left(\frac{Z_1}{2}+Z_2\right)+\frac{Z_1^2}{4}+Z_1Z_2}{Z_2+\frac{Z_1}{2}+Z_R}$$

But $\quad \dfrac{Z_1^2}{4} + Z_1Z_2 = Z_0^2$

$$\therefore \qquad Z_{IN} = \frac{Z_R\left(\dfrac{Z_1}{2} + Z_2\right) + Z_0^2}{Z_2 + \dfrac{Z_1}{2} + Z_R}$$

$$= Z_0 \frac{Z_R\left(1 + \dfrac{Z_1}{2Z_2}\right) + Z_0 \dfrac{Z_0}{Z_2}}{Z_0\left(1 + \dfrac{Z_1}{2Z_2}\right) + Z_R \dfrac{Z_0}{Z_2}}$$

But $\quad \dfrac{Z_0}{Z_2} = \sinh \gamma$ (equation 21)

and $\quad 1 + \dfrac{Z_1}{2Z_2} = \cosh \gamma$ (equation 20).

Thus
$$Z_{IN} = Z_0 \frac{Z_R \cosh \gamma + Z_0 \sinh \gamma}{Z_0 \cosh \gamma + Z_R \sinh \gamma} \tag{25}$$

Input impedance of a T section having a high attenuation

Considering equation 25, and replacing γ by $\alpha + j\beta$, gives :—

$$Z_{IN} = Z_0 \frac{Z_R \cosh(\alpha + j\beta) + Z_0 \sinh(\alpha + j\beta)}{Z_0 \cosh(\alpha + j\beta) + Z_R \sinh(\alpha + j\beta)}$$

$$= Z_0 \frac{Z_R \cosh\alpha \cos\beta + jZ_R \sinh\alpha \sin\beta + Z_0 \sinh\alpha \cos\beta + jZ_0 \cosh\alpha \sin\beta}{Z_0 \cosh\alpha \cos\beta + jZ_0 \sinh\alpha \sin\beta + Z_R \sinh\alpha \cos\beta + jZ_R \cosh\alpha \sin\beta}$$

If α is large :—
$\cosh\alpha = \sinh\alpha$
In this case :—

$$Z_{IN} = Z_0 \frac{Z_R \cos\beta + jZ_R \sin\beta + Z_0 \cos\beta + jZ_0 \sin\beta}{Z_0 \cos\beta + jZ_0 \sin\beta + Z_R \cos\beta + jZ_R \sin\beta}$$
$$= Z_0 \tag{26}$$

Hence if the attenuation (α) is large, the input impedance is equal to Z_0 for all values of phase-shift (β) and for all values of terminating impedance (Z_R).

This result is of fundamental importance, and is applicable to all line transmission networks.

Example.—

A T section (Fig. 576) has an attenuation of 3·45 nepers (30 db) and a purely resistive characteristic impedance of 600 Ω. What is the input impedance (*a*) on open-circuit, (*b*) on short-circuit, and (*c*) when terminated in a 2μF condenser (ω = 5000 radians/sec.) ?

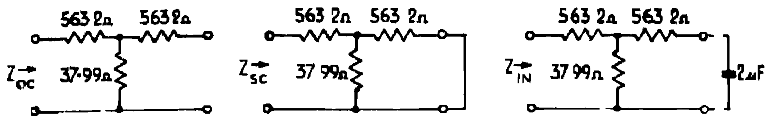


FIG 576

(a) $Z_{OC} = 563\cdot2 + 37\cdot99 = 601$ ohms $\angle 0°$ *Ans.*

(b) $Z_{SC} = 563\cdot2 + \dfrac{37\cdot99 \times 563\cdot2}{601\cdot2} = 599$ ohms $\angle 0°$ *Ans.*

(c) $Z_{IN} = 563\cdot2 + \dfrac{37\cdot99\,(563\cdot2 - j\,100)}{37\cdot99 + 563\cdot2 - j\,100}$

$= 599$ ohms $\angle -0°\,2'$ *Ans.*

THE π SECTION

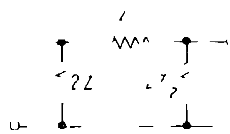
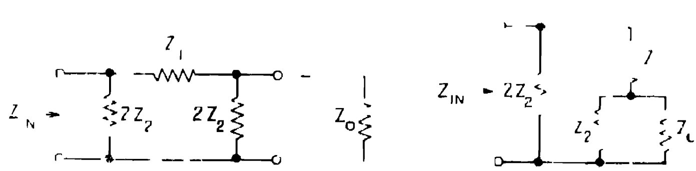


FIG 577—Symmetrical π section

The symmetrical π section, Fig 577, is another very important network encountered in line transmission

Characteristic impedance

The characteristic impedance of the π section may be found in an identical method to that employed for the T section, namely, by terminating the section in Z_0 and equating the input impedance to Z_0



(a) (b)

FIG 578—Symmetrical π section terminated in Z_0

The input impedance Z_{IN} will be $2Z_2$ in parallel with the series combination of Z_1 and ($2Z_2$ and Z_0 in parallel). Fig. 578*b*.

$$\therefore \quad \frac{1}{Z_{IN}} = \frac{1}{2Z_2} + \frac{1}{Z_1 + \dfrac{1}{\dfrac{1}{Z_0} + \dfrac{1}{2Z_2}}} = \frac{1}{Z_1 + \dfrac{2Z_2 Z_0}{2Z_2 + Z_0}}$$

or writing $Y = \dfrac{1}{Z}$ (*i.e.*, using admittances) :—

$$Y_{IN} = \frac{Y_2}{2} + \frac{Y_1\left(Y_0 + \dfrac{Y_2}{2}\right)}{Y_1 + Y_0 + \dfrac{Y_2}{2}}$$

But $Y_{IN} = Y_0$

$$\therefore \quad Y_0 = \frac{Y_2}{2} + \frac{Y_1Y_0 + \dfrac{Y_1Y_2}{2}}{Y_0 + Y_1 + \dfrac{Y_2}{2}}$$

$$\therefore \; Y_0^2 + Y_0Y_1 + \frac{Y_0Y_2}{2} = \frac{Y_0Y_2}{2} + \frac{Y_1Y_2}{2} + \frac{Y_2^2}{4} + Y_1Y_0 + \frac{Y_1Y_2}{2}$$

$$\therefore \quad Y_0^2 = \frac{Y_2^2}{4} + Y_1Y_2$$

$$\therefore \quad Y_0 = \sqrt{\frac{Y_2^2}{4} + Y_1Y_2} \quad \text{(compare this with equation 10)} \qquad (27)$$

This gives $$Z_0 = \frac{1}{\sqrt{\dfrac{1}{4Z_2^2} + \dfrac{1}{Z_1Z_2}}} = \frac{Z_1Z_2}{\sqrt{\dfrac{Z_1^2}{4} + Z_1Z_2}} \qquad (28)$$

Writing Z_{0T} for the characteristic impedance of a T section and $Z_{0\pi}$ for the characteristic impedance of a π section having the same total series and shunt impedances, this gives —

$$Z_{0\pi} = \frac{Z_1Z_2}{Z_{0T}} \qquad (29)$$

Open- and short-circuit impedances

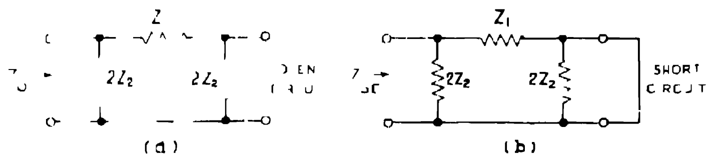


FIG. 579.—Symmetrical π section on open- and short-circuit.

If one side of the π section is open-circuited (*see* Fig. 579*a*) then the input impedance measured at the other side will be :—

$$Z_{OC} = \frac{2Z_2(Z_1 + 2Z_2)}{2Z_2 + Z_1 + 2Z_2} = \frac{2Z_2(Z_1 + 2Z_2)}{Z_1 + 4Z_2} \qquad (30)$$

If one side is short-circuited (*see* Fig. 579*b*), then the input impedance measured at the other side will be :—

$$Z_{SC} = \frac{2Z_1Z_2}{Z_1 + 2Z_2} \qquad (31)$$

The product of these two impedances is :—

$$Z_{oc} \cdot Z_{sc} = \left\{ \frac{2Z_2(Z_1 + 2Z_2)}{Z_1 + 4Z_2} \right\} \times \left\{ \frac{2Z_1Z_2}{(Z_1 + 2Z_2)} \right\}$$

$$= \frac{4Z_1Z_2^2}{Z_1 + 4Z_2}$$

$$= \frac{Z_1^2Z_2^2}{\frac{Z_1^2}{4} + Z_1Z_2}$$

$$= Z_0^2$$

Thus $$Z_0 = \sqrt{Z_{oc} \cdot Z_{sc}} \tag{32}$$

It will be noticed that this result is the same as that obtained for the T section (p 571), *in fact, it is of universal application for any symmetrical section.*

Propagation constant

Consider a π section connected between a generator of internal impedance Z_0 and a load Z_0. Let the input and output currents be I_S and I_R respectively (*see* Fig. 580).

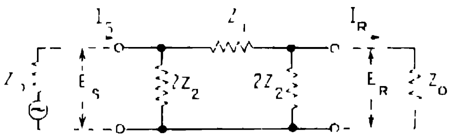


FIG. 580.—Propagation constant of π section.

Since the network is symmetrical, by definition :—

$$\frac{I_S}{I_R} = \frac{E_S}{E_R} = e^{\gamma}$$

From Fig. 580 it will be seen that :

$$E_R = \frac{\dfrac{Z_0 \cdot 2Z_2}{Z_0 + 2Z_2}}{Z_1 + \dfrac{Z_0 \cdot 2Z_2}{Z_0 + 2Z_2}} \cdot E_S$$

$$\therefore \qquad \frac{E_S}{E_R} = e^{\gamma} = 1 + \frac{Z_1(Z_0 + 2Z_2)}{Z_0 \cdot 2Z_2}$$

Let $$Y_0 = \frac{1}{Z_0},\ Y_1 = \frac{1}{Z_1} \text{ and } Y_2 = \frac{1}{Z_2}$$

$$e^{\gamma} = 1 + \frac{1}{Y_1}\left(Y_0 + \frac{Y_2}{2}\right)$$

$$e^{\gamma} = 1 + \frac{Y_2}{2Y_1} + \frac{Y_0}{Y_1} \tag{33}$$

Hence $$\gamma = \log_e \left\{ 1 + \frac{Y_2}{2Y_1} + \frac{Y_0}{Y_1} \right\} \qquad (34)$$

This may be compared with the value of propagation constant obtained for a T section on page 573, it is important to note that it gives the same value

For $$\gamma_\pi = \log_e \left\{ 1 + \frac{Y_2}{2Y_1} + \frac{Y_{0\pi}}{Y_1} \right\}$$

$$= \log_e \left\{ 1 + \frac{Z_1}{2Z_2} + \frac{Z_1}{Z_{0\pi}} \right\}$$

$$= \log_e \left\{ 1 + \frac{Z_1}{2Z_2} + \frac{Z_1 Z_{0T}}{Z_1 Z_2} \right\}$$

i e $$\gamma_\pi = \log_e \left\{ 1 + \frac{Z_1}{2Z_2} + \frac{Z_{0T}}{Z_2} \right\} \qquad (35)$$

$$= \gamma_T$$

Hence all other expressions for γ derived for the T section apply also to the π section—e g , $\cosh \gamma = 1 + \dfrac{Z_1}{2Z_2}$

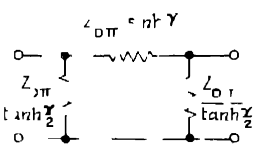


FIG 581 π section showing values of components in terms of characteristic impedance and propagation constant

The π section having characteristic impedance $Z_{0\pi}$ and propagation constant γ is shown in Fig 581

THE HALF-SECTION

Both the symmetrical T section of Fig 582*a* and the symmetrical π section of Fig 582*b* may be split into two half sections It will be seen that the resultant half-sections are identical, as shown in Fig 582*c* , thus either a T section or a π section may be constructed from two such half-sections

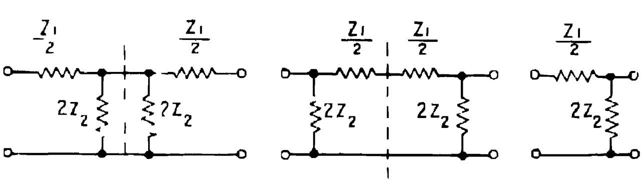


FIG 582 —Showing how (*a*) a T section, and (*b*) a π section, is composed of two half-sections as in (*c*)

A half-section is an example of an asymmetrical network; the expression "characteristic impedance" does not therefore apply, and the network must be considered from the point of view of either iterative impedances or image impedances.

Iterative impedances

The two iterative impedances of a half-section may be found by calculating the input impedance Z'_0 when the section is terminated at the other pair of terminals in Z'_0.

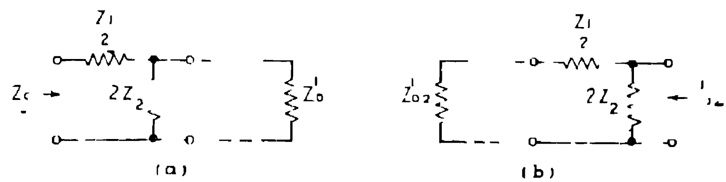


FIG. 583.—Iterative impedances of half section.

Considering the half section shown in Fig. 583*a*:—

$$Z'_{01} = \frac{Z_1}{2} + \frac{2Z_2Z'_{01}}{2Z_2 + Z'_{01}}$$

$$2Z_2Z'_{01} + Z'^2_{01} = Z_1Z_2 + \tfrac{1}{2}Z_1Z'_{01} + 2Z_2Z'_{01}$$

$$\therefore \quad Z'^2_{01} - \tfrac{1}{2}Z_1Z'_{01} - Z_1Z_2 = 0$$

$$Z'_{01} = \sqrt{\frac{Z_1^2}{16} + Z_1Z_2} + \frac{Z_1}{4} \tag{36}$$

Again, considering the half-section shown in Fig. 583*b*:—

$$Z'_{02} = \frac{2Z_2(\tfrac{1}{2}Z_1 + Z'_{02})}{2Z_2 + \tfrac{1}{2}Z_1 + Z'_{02}}$$

$$Z'_{02} = \sqrt{\frac{Z_1^2}{16} + Z_1Z_2} - \frac{Z_1}{4} \tag{37}$$

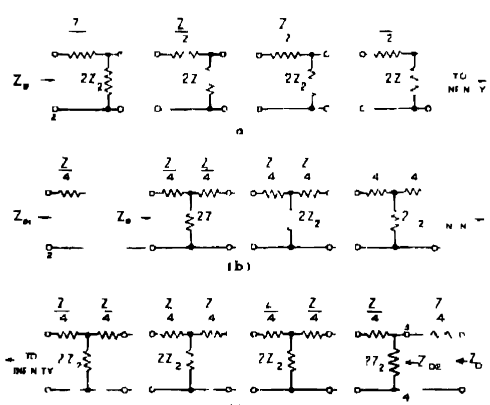


FIG. 584.—Ladder structure formed by an infinite number of half-sections.

The ladder structure formed by an infinite number of half-sections is shown in Fig. 584a; it can be seen to be identical with the series of T sections shown in Fig 584b and c having series-arms $\frac{Z_1}{4}$ and shunt-arms $2Z_2$, with a series element $\frac{Z_1}{4}$ added at the input at 1, 2, and subtracted from the input at 3, 4. The characteristic impedance of the T section shown is $\sqrt{\frac{Z_1^2}{16} + Z_1Z_2}$. The two iterative impedances of the half-sections are thus greater and less than this by $\frac{Z_1}{4}$.

Image impedances

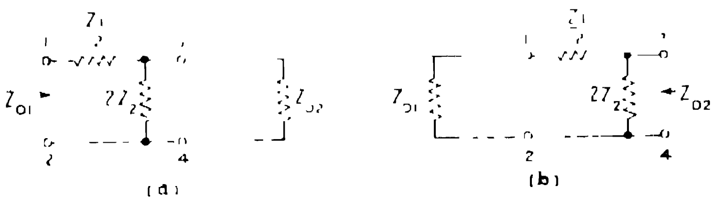


FIG. 585 — Image impedances of half-section.

Let the image impedances of the section be Z_{01} at terminals 1, 2, and Z_{02} at terminals 3, 4 (see Fig. 585).

By definition, if an impedance Z_{02} is connected to terminals 3, 4, the input impedance at 1, 2 will be Z_{01}.

$$Z_{01} = \tfrac{1}{2}Z_1 + \frac{2Z_2Z_{02}}{2Z_2 + Z_{02}} = \frac{Z_1Z_2 + \tfrac{1}{2}Z_1Z_{02} + 2Z_2Z_{02}}{2Z_2 + Z_{02}}$$

$$\therefore\ Z_{01}Z_{02} + 2Z_2Z_{01} - (\tfrac{1}{2}Z_1 + 2Z_2)Z_{02} = Z_1Z_2 \qquad (38)$$

Similarly, if an impedance Z_{01} is connected to terminals 1, 2, the input impedance at 3, 4 will be Z_{02}.

From Fig. 585b:—

$$Z_{02} = \frac{2Z_2\left(\frac{Z_1}{2} + Z_{01}\right)}{2Z_2 + \frac{Z_1}{2} + Z_{01}}$$

$$\therefore\ Z_{01}Z_{02} - 2Z_2Z_{01} + \left(\frac{Z_1}{2} + 2Z_2\right)Z_{02} = Z_1Z_2 \qquad (39$$

Subtracting equation 39 from 38:—

$$4Z_2Z_{01} - (Z_1 + 4Z_2)Z_{02} = 0$$

$$\text{Whence}\qquad \frac{Z_{01}}{Z_{02}} = \frac{\frac{Z_1}{4} + Z_2}{Z_2} \qquad (40)$$

Adding equations 38 and 39:—

$$Z_{01}Z_{02} = Z_1Z_2$$

$$\therefore \qquad Z_{01}^2 = \left(\frac{\frac{Z_1}{4} + Z_2}{Z_2}\right) Z_1Z_2$$

$$\therefore \qquad Z_{01} = \sqrt{\frac{Z_1^2}{4} + Z_1Z_2} = Z_{0T} \tag{41}$$

and

$$Z_{02} = \frac{Z_1Z_2}{\sqrt{\frac{Z_1^2}{4} + Z_1Z_2}} = Z_{0\pi} \tag{42}$$

The image impedances on the two sides of a half-section having a series impedance $\frac{Z_1}{2}$ and a shunt impedance $2Z_2$ are thus seen

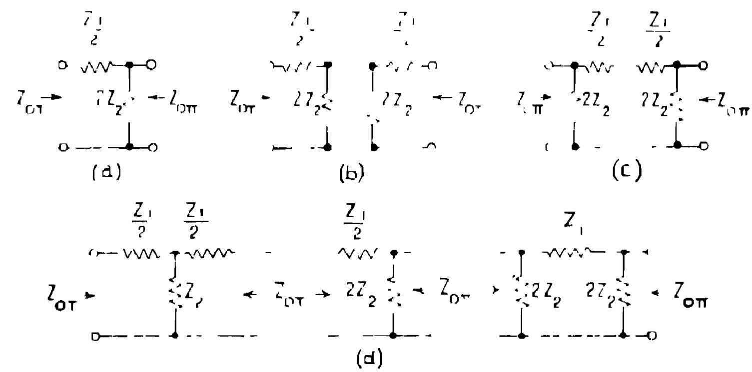


FIG. 586.—Matching a T and a π section, using a half-section.

to be Z_{0T} and $Z_{0\pi}$ (Fig. 586*a*). These impedances are equal respectively to the characteristic impedance of a T section and of a π section having a total series impedance Z_1 and a total shunt impedance Z_2, that is, to the characteristic impedance of that T and of that π section which are produced by combining two such half-sections (Fig. 586*b* and *c*). A half-section can therefore be used for matching between a T section and a π section if both these have the same total series and shunt impedances (*see* Fig. 586*d*).

Open- and short-circuit impedances

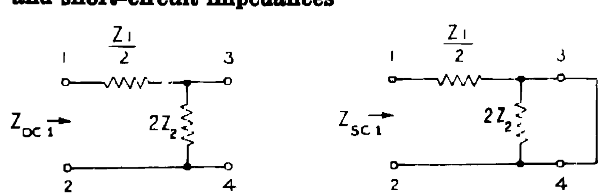


FIG. 587.—Impedances of half-section on open- and short-circuit, looking in at terminals 1 and 2.

The impedances presented at terminals 1, 2 of the half-section (Fig. 587) when the terminals 3, 4 are open- and short-circuited are :—

$$Z_{oc1} = \frac{Z_1}{2} + 2Z_2 \tag{43}$$

$$Z_{sc1} = \frac{Z_1}{2} \tag{44}$$

Hence $\sqrt{Z_{oc1}Z_{sc1}} = \sqrt{\left(\frac{Z_1}{2} + 2Z_2\right)\frac{Z_1}{2}}$

$$= \sqrt{\frac{Z_1^2}{4} + Z_1Z_2} \tag{45}$$

= image impedance looking in at 1, 2

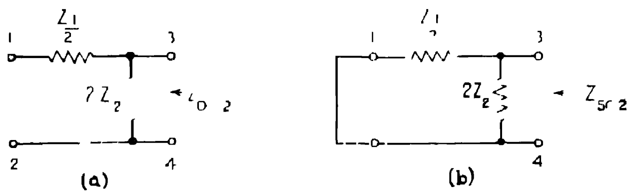


Fig. 588 Impedances of half section on open- and short-circuit, looking in at terminals 3 and 4

Similarly, the impedances presented at terminals 3, 4 when 1, 2 are open and short circuited (Fig. 588) are

$$Z_{oc2} = 2Z_2$$

$$Z_{sc2} = \frac{Z_1Z_2}{\frac{Z_1}{2} + 2Z_2}$$

Hence $\sqrt{Z_{oc2}Z_{sc2}} = \sqrt{\dfrac{2Z_1Z_2^2}{\frac{Z_1}{2} + 2Z_2}}$

$$= \frac{Z_1Z_2}{\sqrt{\frac{Z_1^2}{4} + Z_1Z_2}} \tag{46}$$

= image impedance looking in at 3, 4

Thus the impedances obtained by taking the geometric mean of the open and short-circuit impedances at the two sides of the half-section are seen to be equal to the image impedance *This is true for any asymmetrical network and provides a convenient method for the determination of the image impedances*

THE L SECTION

The ladder network of Fig 562*a* and *b* can be analysed into a series of " L " or " C " sections (Fig. 589) instead of T or π sections

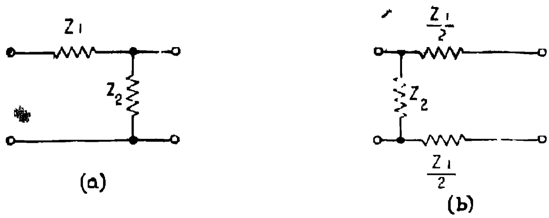


FIG. 589 —Unbalanced and balanced forms of L section

Such sections are frequently used for matching purposes.

Iterative impedances

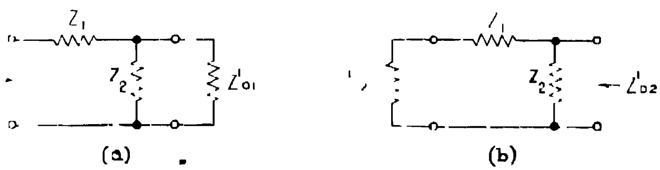


FIG. 590 —Iterative impedances of L section.

The iterative impedances of the L section (*see* Fig. 590) may be found by the same method as adopted for the half-section :—

$$Z'_{01} = \sqrt{\frac{Z_1^2}{4} + Z_1Z_2} + \frac{Z_1}{2} \tag{47}$$

$$Z'_{02} = \sqrt{\frac{Z_1^2}{4} + Z_1Z_2} - \frac{Z_1}{2} \tag{48}$$

Image impedances

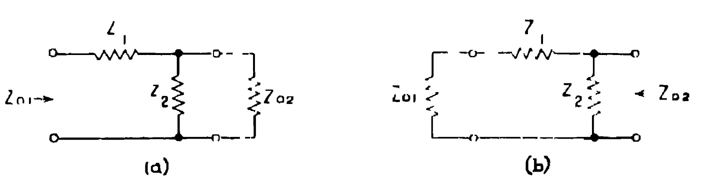


FIG. 591.—Image impedances of L section.

The image impedances of the L section (*see* Fig. 591) may be found by the method previously adopted ; they are of course also equal to the geometric means of the corresponding open- and short-circuit impedances.

$$Z_{01} = \sqrt{Z_{oc1} \cdot Z_{sc1}} = \sqrt{Z_1Z_2 + Z_1^2} \tag{49}$$

$$Z_{02} = \sqrt{Z_{oc2} \cdot Z_{sc2}} = \frac{Z_1Z_2}{\sqrt{Z_1Z_2 + Z_1^2}} \tag{50}$$

Z_1 and Z_2 may be obtained in terms of Z_{01} and Z_{02}, giving :—

$$Z_1 = \sqrt{Z_{01}(Z_{01} - Z_{02})} \qquad (51)$$

$$Z_2 = Z_{02}\sqrt{\frac{Z_{01}}{Z_{01} - Z_{02}}} \qquad (52)$$

THE LATTICE SECTION

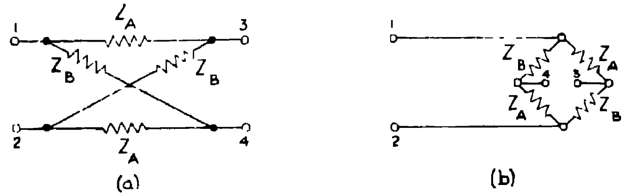


FIG. 592.—Lattice section, with representation as a bridge.

Lattice sections are symmetrical and balanced, and will therefore have characteristic impedances and propagation constants. The lattice structure, Fig. 592*a*, may be redrawn as a bridge structure, Fig. 592*b*.

Characteristic impedance

Let the lattice section of Fig. 593*a* be terminated at terminals 3, 4 in an impedance Z_0. Then if Z_0 is the characteristic impedance of the section, the impedance looking into terminals 1, 2 will also be Z_0.

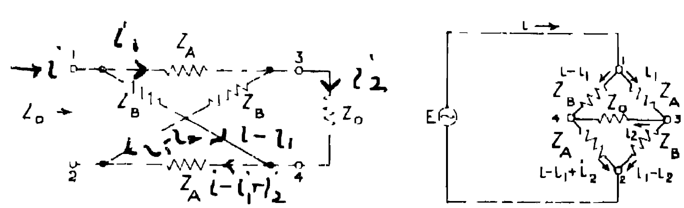


FIG. 593.—Characteristic impedance of lattice section

The value of impedance looking into terminals 1, 2 can be found by the application of Kirchhoff's Laws. The meshes may be seen more clearly by redrawing Fig. 593*a* in the bridge form, as in Fig. 593*b*.

Let the voltage applied to terminals 1, 2, be E, and let currents i, and i_2 flow as indicated.

From the mesh 1, 4, 3, 2 :—

$$E = (i - i_1)Z_B - i_2Z_0 + (i_1 - i_2)Z_B$$

$$E = iZ_B - i_2(Z_0 + Z_B) \qquad (53)$$

From the mesh 1, 3, 4, 2 :—

$$E = i_1 Z_A + i_2 Z_0 + (i - i_1 + i_2) Z_A$$

i.e.
$$E = iZ_A + i_2(Z_0 + Z_A) \tag{54}$$

From these two equations :—

$$i_2 = \frac{E - iZ_A}{Z_0 + Z_A} = \frac{iZ_B - E}{Z_0 + Z_B} \tag{55}$$

$$\therefore \frac{E}{i}(2Z_0 + Z_A + Z_B) = Z_A(Z_0 + Z_B) + Z_B(Z_0 + Z_A)$$

But $\frac{E}{i}$ = input impedance at 1, 2, $= Z_0$

$$\therefore \quad Z_0(2Z_0 + Z_A + Z_B) = Z_A(Z_0 + Z_B) + Z_B(Z_0 + Z_A)$$

$$\therefore \quad 2Z_0^2 = 2Z_A Z_B$$

$$Z_0 = \sqrt{Z_A Z_B} \tag{56}$$

Open- and short-circuit impedances

The impedances presented at one pair of terminals of a lattice section, when the other pair is open- and short-circuited, are respectively :—

$$Z_{OC} = \tfrac{1}{2}(Z_A + Z_B) \tag{57}$$

$$Z_{SC} = \frac{2Z_A Z_B}{Z_A + Z_B} \tag{58}$$

$$\therefore \quad \sqrt{Z_{OC} \cdot Z_{SC}} = \sqrt{\tfrac{1}{2}(Z_A + Z_B) \cdot \frac{2Z_A Z_B}{Z_A + Z_B}} = \sqrt{Z_A Z_B} = Z_0 \tag{59}$$

This verifies once again that the geometric mean of the open- and short-circuit impedances equals the characteristic impedance Z_0 (image impedance in the case of asymmetrical sections).

Propagation constant

Let γ be the propagation constant. By definition :—

$$\frac{i}{i_2} = e^{\gamma}$$

From equation 55 above :—

$$i_2 = \frac{E - iZ_A}{Z_0 + Z_A} = \frac{iZ_B - E}{Z_0 + Z_B}$$

But $\quad E = iZ_0$

$$\therefore \quad i_2 = i\frac{Z_0 - Z_A}{Z_0 + Z_A} = i\frac{Z_B - Z_0}{Z_0 + Z}$$

Hence
$$e^{\gamma} = \frac{Z_0 + Z_A}{Z_0 - Z_A} = \frac{Z_B + Z_0}{Z_B - Z_0} \tag{60}$$

or
$$\gamma = \log_e \frac{Z_0 + Z_A}{Z_0 - Z_A} = \log_e \frac{Z_B + Z_0}{Z_B - Z_0} \tag{61}$$

In addition, since $e^{\gamma} = \dfrac{Z_0 + Z_A}{Z_0 - Z_A}$,

$$Z_A\,(e^{\gamma} + 1) = Z_0(e^{\gamma} - 1)$$

$$\therefore \qquad \frac{Z_A}{Z_0} = \frac{e^{\gamma} - 1}{e^{\gamma} + 1} = \frac{e^{\frac{\gamma}{2}} - e^{-\frac{\gamma}{2}}}{e^{\frac{\gamma}{2}} + e^{-\frac{\gamma}{2}}} = \tanh\frac{\gamma}{2} \tag{62}$$

$$\therefore \qquad \tanh\frac{\gamma}{2} = \frac{Z_A}{Z_0} = \frac{Z_A}{\sqrt{Z_A Z_B}} = \sqrt{\frac{Z_A}{Z_B}} \tag{63}$$

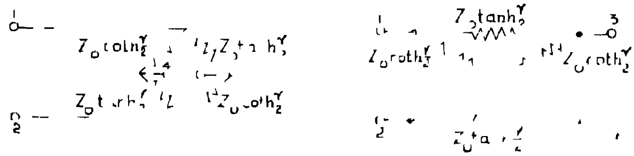


FIG 594.—Lattice section, showing values of components in terms of characteristic impedance and propagation constant.

The lattice section having characteristic impedance Z_0 and propagation constant γ, deduced from equations 56 and 63, is shown in Fig 594.

THE BRIDGED-T SECTION

Characteristic impedance

The characteristic impedance of the bridged-T section may be found, as in the case of the lattice section, by assuming one pair of terminals to be terminated in Z_0 and an EMF E to be applied

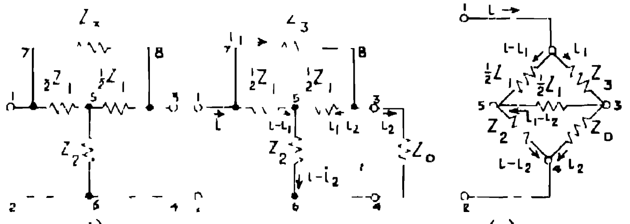


FIG. 595.—Bridged-T section.

to the other pair of terminals. Then, if currents flow as indicated in Fig 595*b*, the following equations are obtained :—

From the mesh 178342 :— $i_1 Z_3 + i_2 Z_0 = E$ (64)

$$\text{whence } i_1 = \frac{E}{Z_3} - \frac{i_2 Z_0}{Z_3}$$

From the mesh 15342 :— $i\,.\,\dfrac{Z_1}{2} - i_1 Z_1 + i_2\left(\dfrac{Z_1}{2} + Z_0\right) = E$ (65)

From the mesh 1562 :— $i\left(\frac{Z_1}{2} + Z_2\right) - i_1\frac{Z_1}{2} - i_2 Z_2 = E$ (66)

Solving these equations gives :—

$$Z_0 = \sqrt{\left(\frac{Z_1^2}{4} + Z_1 Z_2\right)\frac{Z_3}{Z_1 + Z_3}} \qquad (67)$$

In practice, the two series links $\frac{Z_1}{2}$ are frequently made equal to Z_0. Equation 67 then reduces to :—

$$Z_0 = \sqrt{Z_2 Z_3} \qquad (68)$$

Open- and short-circuit impedances of bridged-T section

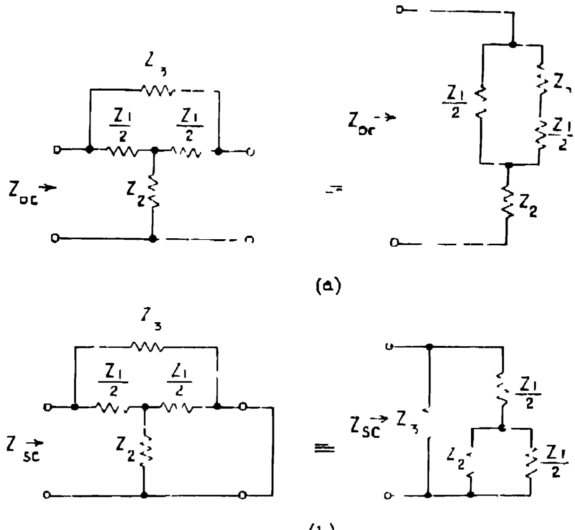


FIG 596.— Bridged-T section on open- and short-circuit

The impedances presented at one pair of terminals of the bridged-T section, while the other pair of terminals is open-circuited and short-circuited, are respectively :—

$$\begin{aligned} Z_{oc} &= Z_2 + \frac{\frac{1}{2}Z_1(\frac{1}{2}Z_1 + Z_3)}{Z_1 + Z_3} \\ &= \frac{Z_2(Z_1 + Z_3) + \frac{1}{2}Z_1(\frac{1}{2}Z_1 + Z_3)}{Z_1 + Z_3} \end{aligned} \qquad (69)$$

$$Z_{sc} = \frac{Z_3\left\{\frac{1}{2}Z_1 + \frac{\frac{1}{2}Z_1 Z_2}{\frac{1}{2}Z_1 + Z_2}\right\}}{Z_3 + \frac{1}{2}Z_1 + \frac{\frac{1}{2}Z_1 Z_2}{\frac{1}{2}Z_1 + Z_2}}$$

$$= \frac{Z_1 Z_3(\frac{1}{4}Z_1 + Z_2)}{\frac{1}{2}Z_1(\frac{1}{2}Z_1 + Z_3) + Z_2(Z_1 + Z_3)}$$

$$\therefore \quad \sqrt{Z_{oc} Z_{sc}} = \sqrt{\frac{Z_1 Z_3(\frac{1}{4}Z_1 + Z_2)}{Z_1 + Z_3}}$$

$$= \sqrt{\left(\frac{Z_1^2}{4} + Z_1 Z_2\right)\frac{Z_3}{Z_1 + Z_3}}$$

Hence $\qquad Z_0 = \sqrt{Z_{oc} Z_{sc}}$

Propagation constant

From equations 64 and 65 —

$$i_2 = \frac{i_1 Z_0(Z_1 + Z_3) - i_1 \frac{1}{2} Z_1 Z_3}{Z_0 Z_1 + \frac{1}{2} Z_1 Z_3 + Z_3 Z_0}$$

$$= \frac{i_1\{Z_0(Z_1 + Z_3) - \frac{1}{2} Z_1 Z_3\}}{Z_0(Z_1 + Z_3) + \frac{1}{2} Z_1 Z_3}$$

$$\therefore \quad \frac{i_1}{i_2} = \frac{Z_0(Z_1 + Z_3) + \frac{1}{2} Z_1 Z_3}{Z_0(Z_1 + Z_3) - \frac{1}{2} Z_1 Z_3}$$

$$\therefore \quad \gamma = \log_e \frac{Z_0(Z_1 + Z_3) + \frac{1}{2} Z_1 Z_3}{Z_0(Z_1 + Z_3) - \frac{1}{2} Z_1 Z_3} \tag{71}$$

When $\frac{Z_1}{2} = Z_0$, this reduces to

$$\gamma = \log_e\left(1 + \frac{Z_3}{Z_0}\right) \tag{72}$$

or, on substituting for Z_3 from equation 68, to

$$\gamma = \log_e\left(1 + \frac{Z_0}{Z_2}\right) \tag{73}$$

NETWORK EQUIVALENCE THEOREMS

Equivalence of T and π sections*

Theorem. At any one frequency a T section can be interchanged, in any network, with a π section, and *vice versa*, provided that certain relations are maintained between the elements of the two sections.

Since the T section, Fig. 597*a*, may be redrawn as a *star* (Fig. 597*b*), and the π section, Fig. 597*c*, may be redrawn as a *mesh* (Fig. 597*d*), this is sometimes known as a "star-mesh conversion".

If the impedance looking into the terminals 1 and 3 are equated for the two sections, the following relationship is obtained. —

$$Z_1 + Z_2 = \frac{Z_C(Z_A + Z_B)}{Z_A + Z_B + Z_C} \tag{74}$$

* The T and π sections so far discussed have *not* been equivalent, but have been related to the same ladder network. In this section electrically equivalent T and π networks are considered.

Similarly, equating impedances looking into terminals 3 and 4,

$$Z_2 + Z_3 = \frac{Z_A(Z_B + Z_C)}{Z_A + Z_B + Z_C} \tag{75}$$

and equating impedances looking into terminals 1 and 2 —

$$Z_3 + Z_1 = \frac{Z_B(Z_C + Z_A)}{Z_A + Z_B + Z_C} \tag{76}$$

Now add equations 74 and 76 and subtract 75,

$$Z_1 = \frac{Z_B Z_C}{Z_A + Z_B + Z_C} = \frac{Z_B Z_C}{\Delta} \tag{77}$$

where $\Delta = Z_A + Z_B + Z_C$

Similarly, adding 74 and 75 and subtracting 76,

$$Z_2 = \frac{Z_C Z_A}{Z_A + Z_B + Z_C} = \frac{Z_C Z_A}{\Delta} \tag{78}$$

Finally adding 75 and 76 and subtracting 74

$$Z_3 = \frac{Z_A Z_B}{Z_A + Z_B + Z_C} = \frac{Z_A Z_B}{\Delta} \tag{79}$$

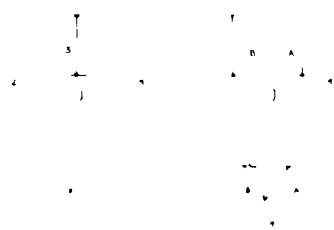


FIG. 597 — T section represented as a star and π section represented as a mesh.

These equations, 77, 78 and 79, give the relations between the impedance elements for a certain impedance equivalence between the two sections under rather special conditions. It will now be shown that these equations give the equivalence of the two sections under the more stringent conditions depicted in Fig. 598a and b. In Fig. 598a the T section is used to connect a

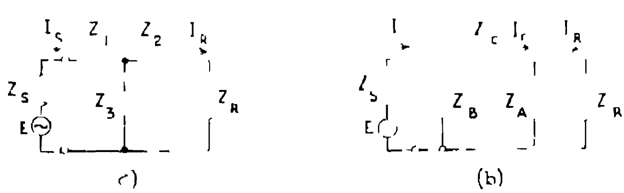


FIG. 598 — (a) T section, and (b) π section, connecting a generator to a load.

generator of internal impedance Z_S and voltage E to a lo impedance Z_R. In Fig. 598b the π section is used to connect th same generator to the same load. The equivalence of the tw sections will be proved by showing that the currents I_S and I_R a the same for the two networks.

From Fig. 598a the current I_S is seen to be

$$I_S = \frac{E}{Z_S + Z_1 + \dfrac{Z_3(Z_2 + Z_R)}{Z_2 + Z_3 + Z_R}} \qquad (80)$$

Now suppose that Z_1, Z_2 and Z_3 have the values given by equations 77, 78 and 79 respectively. Substituting these values in equation 80 gives

$$I_S = \frac{E}{Z_S + \dfrac{Z_B(Z_R Z_C \quad Z_1 Z_A \quad Z_C Z_A)}{Z_C Z_A + Z_1 Z_B \quad Z\, Z_R}} \qquad (81)$$

From Fig. 598b the current I'_S is seen to be

$$I'_S = \frac{E}{Z_S + \dfrac{Z_B\left(Z_C + \dfrac{Z_1 Z_R}{Z_A + Z_R}\right)}{Z_B \quad Z + \dfrac{Z_1 Z_R}{Z_A + Z_R}}}$$

i.e.

$$I'_S = \frac{E}{Z \quad \dfrac{Z_B Z_C Z_A \quad Z_R Z_C \quad Z_R Z_A)}{Z_1 Z_B + Z_C Z_A + Z_R}} \qquad (82)$$

Hence $I_S = I'_S$

From Fig. 598a

$$I_R = \frac{Z_3}{Z \quad Z \quad Z_3} I_S$$

or if Z_1, Z_2 and Z_3 have the values assigned to them by equations 77, 78 and 79 respectively

$$I_R = \frac{Z_1 Z_B}{Z Z_1 + Z_1 Z_A} \quad Z_A\, I_S \qquad (83)$$

From Fig. 598b

$I_R = \dfrac{Z_A}{Z_1 + Z_R} I$ where I_C is the current in the impedance Z.

Similarly

$$I = \frac{Z_A}{Z_B + Z_C + \dfrac{Z_A Z_R}{Z_A + Z_R}} I_S$$

$$I_R = \frac{Z_A Z_B}{Z_A Z_B + Z_C Z_A + \Delta Z_R} I'_S \qquad (84)$$

But $I'_S = I_S$, hence, comparing equations 83 and 84, it will be seen that $I'_R = I_R$.

The two sections are therefore equivalent when a generator of impedance Z_S is connected to terminals 1 and 3 and a load impedance Z_R to terminals 2 and 3. Since the equations 81, 82, 83 and 84 are all symmetrical with respect to Z_A, Z_B and Z_C (that is, they are unchanged if Z_A, Z_B and Z_C are interchanged in cyclic order), it follows that the sections are equivalent if the load and generator are each connected to any two of the three terminals. Finally, since E, Z_S and Z_R are quite general values, the equivalence in all respects of the two sections is established.

Given a π section, therefore, it is possible to replace it in any network by the equivalent T section whose elements are given by equations 77, 78 and 79.

Let $\Delta' = Z_2Z_3 + Z_3Z_1 + Z_1Z_2$

Then from equations 77, 78 and 79 —

$$\Delta' = \frac{Z_A^2Z_BZ_C + Z_B^2Z_CZ_A + Z_C^2Z_AZ_B}{\Delta^2}$$

i.e. $\Delta' = \dfrac{Z_AZ_BZ_C}{\Delta}$

i.e. $\dfrac{\Delta'}{Z_1} = Z_A$, using equation 77

i.e. $$Z_A = \frac{Z_2Z_3 + Z_3Z_1 + Z_1Z_2}{Z_1} \qquad (85)$$

Similarly, $$Z_B = \frac{Z_2Z_3 + Z_3Z_1 + Z_1Z_2}{Z_2} \qquad (86)$$

and $$Z_C = \frac{Z_2Z_3 + Z_3Z_1 + Z_1Z_2}{Z_3} \qquad (87)$$

Therefore, given a T section, it is possible to replace it in any network by the equivalent π section whose elements are given by equations 85, 86 and 87.

It is possible to simplify these results by considering admittances instead of impedances. Thus —

$$\frac{\Delta'}{Z_1Z_2Z_3} = \frac{1}{Z_1} + \frac{1}{Z_2} + \frac{1}{Z_3}$$

i.e. $$\Delta' = \frac{Y_1 + Y_2 + Y_3}{Y_1Y_2Y_3}$$

Now, from (85) —

$$Y_A = \frac{1}{Y_1\Delta'} = \frac{Y_1Y_2Y_3}{Y_1(Y_1 + Y_2 + Y_3)}$$

i.e. $$Y_A = \frac{Y_2Y_3}{Y_1 + Y_2 + Y_3} = \frac{Y_2Y_3}{\Delta''} \qquad (88)$$

where $\Delta'' = Y_1 + Y_2 + Y_3$

Similarly, $$Y_B = \frac{Y_3Y_1}{Y_1+Y_2+Y_3} = \frac{Y_3Y_1}{\Delta''} \qquad (89)$$

and $$Y_C = \frac{Y_1Y_2}{Y_1+Y_2+Y_3} = \frac{Y_1Y_2}{\Delta''} \qquad (90)$$

The significance of the statement that this theorem is true only at a single frequency will be made clear by an example.

Example.—

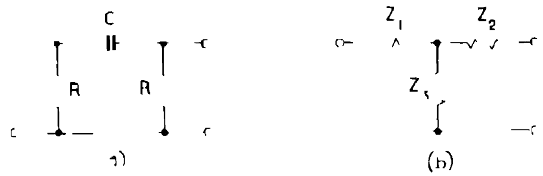


Fig. 599

Suppose the simple π structure of Fig. 599*a* is part of a more complex network and consider a particular frequency having corresponding angular velocity ω_0. Using the notation of Fig. 598*b*,

$$Z_A = Z_B = R \text{ and } Z_C = -\frac{j}{\omega_0 C}$$

Now apply the theorem and replace the π section by the equivalent T section of Fig. 599*b*. Using equations 77 and 79,

$$Z_1 = Z_2 = \frac{-\dfrac{jR}{\omega_0 C}}{2R - \dfrac{j}{\omega_0 C}} = \frac{-jR(2\omega_0 CR + j)}{4\omega_0^2C^2R^2+1}$$

i.e. $$Z_1 = Z_2 = \frac{R}{4\omega_0^2C^2R^2+1} - \frac{j2\omega_0 CR^2}{4\omega_0^2C^2R^2+1}$$

and $$Z_3 = \frac{R^2}{2R - \dfrac{j}{\omega_0 C}} = \frac{(2\omega_0 CR^3 + jR^2)\omega_0 C}{4\omega_0^2C^2R^2+1}$$

Z_1 and Z_2 are equivalent to resistances in series with condensers but the values of resistance and capacity both depend on ω_0. Similarly, Z_3 is equivalent to a resistance in series with an inductance, the values of both components depending on ω_0. If therefore some other frequency is chosen an entirely different T section will result. It is also important to notice that in some cases a resistive or reactive element in the equivalent circuit may turn out to be negative, that is to say, the equivalent circuit may not be physically realizable using passive components.

Equivalence of a complex network to a simple three-element configuration

Theorem.—Any four-terminal network made up of linear

impedances, no matter how complex it may be, can be represented, at a single frequency, by a simple T or π section.

This theorem follows at once from the equivalence of the T and π section. A complicated network can be reduced to a single section by successive transformations from T to π and the reverse, as will be seen from the following example.

Example.—

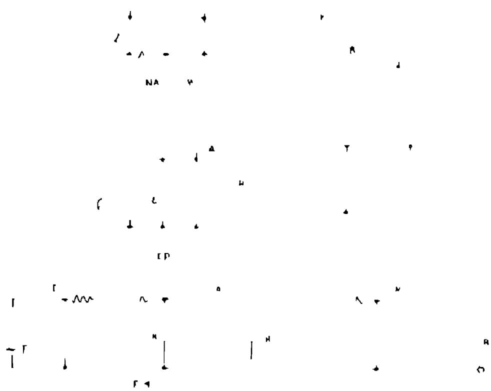


FIG. 600. Reduction of a complex network to a simple T section.

Fig. 600 shows the reduction of a fairly complex network to a simple T section in 5 steps; the number of steps required in a particular case will, of course, increase with the complexity of the network.

Step 1. Reduce the π network consisting of Z_3, Z_4 and Z_6 to a T network Z_A, Z_B, Z_C.

Step 2. Reduce the π network Z_2, Z_1, Z_5 to a T network Z_D, Z_E, Z_F.

Step 3. Reduce the π network Z_8, Z_B, $Z_C + Z_9$ to a T network Z_G, Z_H, Z_I.

Step 4. Transfer the impedance Z_{10} from the lower to the upper arm of the network. This is permissible because the same current flows through Z_{10} and through Z_F and Z_G in series. Z_{10} can therefore be placed in series with Z_F and Z_G in the same arm without affecting the voltage developed across the load impedance Z_R.

Step 5. Reduce the π network $Z_7 + Z_E$, $Z_F + Z_{10} + Z_G$, Z_H to a T network Z_L, Z_M, Z_N.

A simple T section now connects the generator and load, the T section being the equivalent of the original transmission network.

Though it is sometimes useful to perform the actual reduction of a complex network to a simple equivalent T or π section, the more important aspect of this theorem is the fact that it demonstrates the existence of such simple equivalent circuits.

T and π sections equivalent to a perfect transformer

For some purposes it is convenient to replace the separate primary and secondary circuits of a perfect transformer by a circuit linked by a network of impedances. "T" and "π" sections can be used to represent a perfect transformer that has an impedance in series or in shunt with either winding.

The values of the components of these equivalent networks are derived by considering the impedances at the input and at the output terminals with the other terminals open- and short-circuited.

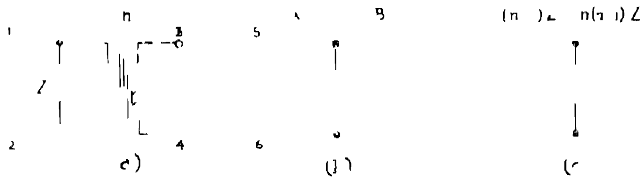


FIG. 601. Transformer with an impedance Z across primary and equivalent T section.

For example suppose that the transformer in Fig. 601a can be replaced by the T network of Fig. 601b. If this supposition is justified then the impedance between terminals 1 and 2 with terminals 3 and 4 open circuited (which will be denoted by Z_{12oc}) must be the same as that (Z_{56oc}) between terminals 5 and 6, with 7 and 8 open circuited.

Hence $\quad Z_{12oc} = Z_{56oc} \quad$ or $\quad Z = A + C \qquad (91)$

Similarly, $\quad Z_{34oc} = Z_{78oc} \quad$ or $\quad n^2Z = B + C \qquad (92)$

and $\quad Z_{12sc} = Z_{56sc} \quad$ or $\quad 0 = A + \dfrac{BC}{B+C} \qquad (93)$

and $\quad Z_{34sc} = Z_{78sc} \quad$ or $\quad 0 = B + \dfrac{AC}{A+C} \qquad (94)$

From (91) and (92) $\quad A = (Z - C) \quad$ and $\quad B = (n^2Z - C)$

From (93), $\quad AB + BC + CA = 0$

$$(Z - C)(n^2Z - C) + (n^2Z - C)C + C(Z - C) = 0$$

$$n^2Z^2 - CZ - n^2CZ + C^2 + n^2CZ - C^2 + CZ - C^2 = 0$$

$$n^2Z^2 - C^2 = 0$$

$$\therefore \qquad C = nZ$$

$$A = Z - nZ = -(n - 1)Z \qquad (95)$$

$$B = n^2Z - nZ = n(n - 1)Z \qquad (96)$$

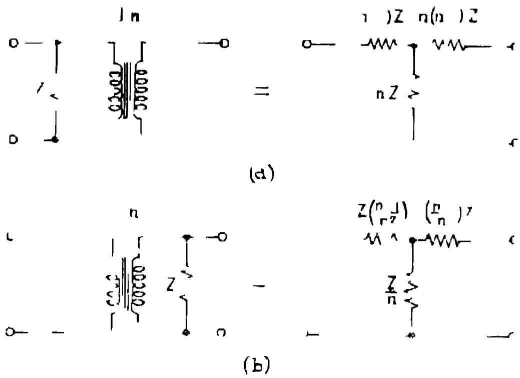


Fig. 602. Transformers with impedances in shunt with primary and secondary windings, together with equivalent T section

The components of the equivalent T network therefore have the values shown in Fig. 601c. Some of the more useful equivalent T and π networks are shown in Figs. 602 and 603. It must be noted that in each case one component is a negative impedance and the equivalent circuit of the transformer alone can never be physically realised using passive components (unless Z is a pure reactance, in which case the equivalent circuit can be physically realised at any one frequency). However, when the transformer is connected into a circuit the negative impedance may come in series or parallel with a larger positive impedance so that the two can be replaced by a single (smaller) positive impedance. The circuit as a whole can then be represented by a physically realisable equivalent circuit.

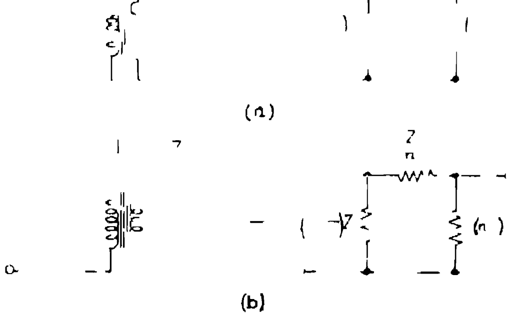


Fig. 603. Transformers with impedances in series with primary and secondary windings, together with equivalent π sections

Development of lattice sections with series impedances

Theorem If any impedance Z be subtracted simultaneously from all four arms of a lattice section, and placed in series with the input and output terminals of the section then the resulting section is electrically identical to the original lattice section

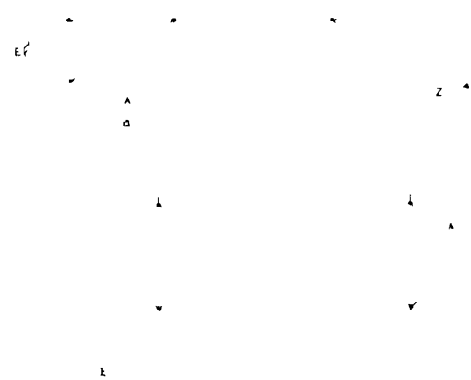


FIG. 604 —Development of lattice section with series impedances

Consider the lattice section shown in Fig. 604*a* and the resulting section (Fig. 604*c*) when it is modified in this way. Let these two sections be terminated on one side with any impedance Z_R and let a voltage E be applied to the other pair of terminals. Then currents flow as indicated in the figure, the application of Kirchhoff's second law to the two sections yields the following equations

Original lattice section (Fig. 604*b*)

Mesh 1432 $\quad E = (i - i_1)Z_I - i_2Z_I - (i_1 - i)Z_B$

$\quad E = iZ_B - i_2(Z_B + Z_R)$ (9

Mesh 1342 $\quad E = i_1Z_A - iZ_R + (i_1 + i_2)Z_A$

$\quad E = iZ_1 - i(Z_A + Z_I)$ (98

Putting Z_R equal to the characteristic impedance Z_0 the equations can be solved as on page 587 to give the characteristic impedance Z_0 and the propagation constant γ namely

$$Z_0 = \sqrt{Z_1Z_I}$$

and $$\log_e \frac{Z_0 + Z_A}{Z_0 - Z_A} \quad \log_e \frac{Z_B - Z_0}{Z_B - Z_0}$$

Modified lattice section (Fig. 604*d*)

Mesh 1432 $\quad E = iZ_I + (i - i_1)(Z_B - Z) - i_2(Z + Z_R) + (i_1 - i)(Z_B - Z) + \frac{1}{2}Z$

$\quad E = iZ_B - i_2(Z_L + Z_B)$ (99

Mesh 1342 —
$$E = \tfrac{1}{2}Zi + i_1(Z_A - Z) + i_2(Z + Z_B) + (i - i_1 + i_2)(Z_A - Z) + \tfrac{1}{2}Zi$$

i.e.
$$E = iZ_A + i_2(Z_B + Z_A) \qquad (100)$$

Equations 99 and 100 are seen to be identical with equations 97 and 98 respectively, and they will therefore yield the same results for Z_0 and for γ. The modified lattice section of Fig. 604c is therefore electrically equivalent to the original lattice section of Fig. 604a.

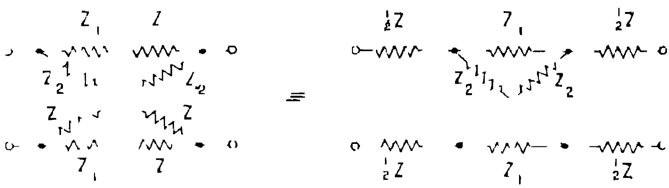


Fig. 605.— Lattice section, and equivalent developed section with series impedances.

In practice, this theorem is most useful when all four arms of the lattice section have a common series impedance. This is illustrated in Fig. 605, where $Z_A = Z_1 + Z$ and $Z_B = Z_2 + Z$.

Equivalence of a lattice to a T section

Theorem. A lattice section can be interchanged, in any network, with a T section, and *vice versa*, provided that certain relations are maintained between the elements of the two sections.

This theorem follows from the last theorem. No restriction was imposed on the value of the impedance Z that was subtracted

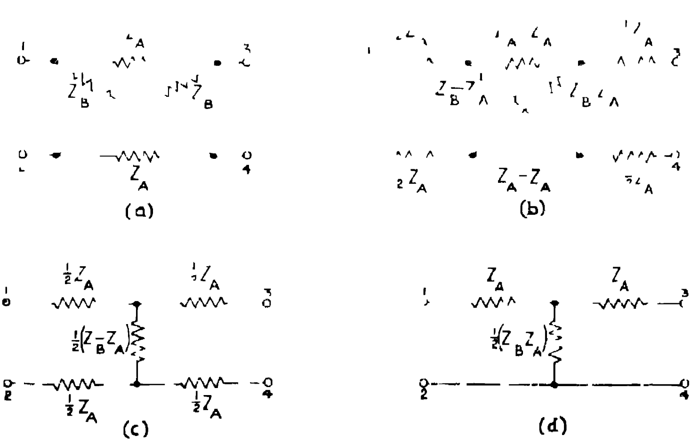


Fig. 606.— Equivalence of lattice and T sections.

from all four arms of the lattice section, and placed in series with the terminals ; Z can therefore be made equal to Z_A.

When Z_A is subtracted from all four arms of the section shown in Fig. 606a, and placed in series with the terminals, the "lattice arms are seen (Fig. 606b) to be in parallel, and the resulting section can be redrawn as in Fig. 606c. This is a balanced T section, the unbalanced form of which is shown in Fig. 606d.

The characteristic impedance and propagation constant of the T section shown in Fig. 606d are given by : –

$$Z_{0T} = \sqrt{\tfrac{1}{4}(2Z_A)^2 + 2Z_A\tfrac{1}{2}(Z_B - Z_A)}$$
$$= \sqrt{Z_A Z_B} \qquad (10\ldots$$

and
$$\gamma_T = \log_e \left\{1 + \frac{2Z_A}{(Z_B - Z_A)} + \frac{2Z_0}{(Z_B - Z_A)}\right\}$$
$$= \log \left(\frac{Z_A + 2Z_0 + Z_B}{Z_B - Z_A}\right)$$
$$= \log_e \frac{\sqrt{Z_B} + \sqrt{Z_A}}{\sqrt{Z_B} - \sqrt{Z_A}} = \log_e \frac{Z_0 + Z_A}{Z_0 - Z_A} \qquad (1\ldots$$

These results are identical with those obtained on page 587 for the original lattice section of Fig. 606a, and therefore the two sections are electrically equivalent.

Development of lattice sections with shunt impedances

Theorem. If any admittance Y be subtracted simultaneously from all four arms of a lattice section, and placed across the input and output terminals, then the resulting section is electrically identical to the original lattice section.

Consider the lattice section shown in Fig. 607a, and the resulting section (Fig. 607d) when it is modified in this way. Let these two sections be terminated on one side with any admittance Y_R, and let a voltage E, applied to the other pair of terminals, cause an input current i. Let the voltages appearing across the various components of the two networks be as shown in Figs. 607b and e respectively; then the currents flowing will be as indicated in Figs. 607c and and the application of Kirchhoff's first law to each terminal in turn yields the following equations :

Original lattice section (Fig. 607c) : –

Terminal 1. $\quad i = Y_B(E_1 + E_2) + Y_A E_1$

i.e. $\quad i = E_1(Y_A + Y_B) + E_2 Y_B \qquad (10\ldots$

Terminal 2.– $\quad i = Y_A(E - E_1 - E_2) + Y_B(E - E_1)$

i.e. $\quad i = E(Y_A + Y_B) - E_1(Y_A + Y_B) - E_2 Y_A \qquad (10\ldots$

Terminal 3.— $\quad 0 = Y_B(E - E_1) + Y_B E_2 - Y_A E_1$

i.e. $\quad 0 = EY_B - E_1(Y_A + Y_B) + E_2 Y_B \qquad (10\ldots$

From (103) and (105)

$$i = EY_B + E_2(Y_B + Y_l) \tag{106}$$

From (104) and (105)

$$i = EY_A - E_2(Y_A + Y_l) \tag{107}$$

Putting Y_l equal to Y_0 and eliminating E_2 from equations 106 and 107 one obtains

$$Y_0 = \frac{i}{E} = \frac{Y_A(Y_B + Y_0) + Y_B(Y_A + Y_0)}{Y_A + Y_B + 2Y_0}$$

Whence $\qquad Y_0 = \sqrt{Y_A Y_B}$

And therefore $\qquad Z_0 = Y_0^{-1} = \dfrac{1}{\sqrt{Y_A Y_B}}$

$$= \sqrt{Z_A Z_B} \qquad \text{(as seen on p. 587)} \tag{108}$$

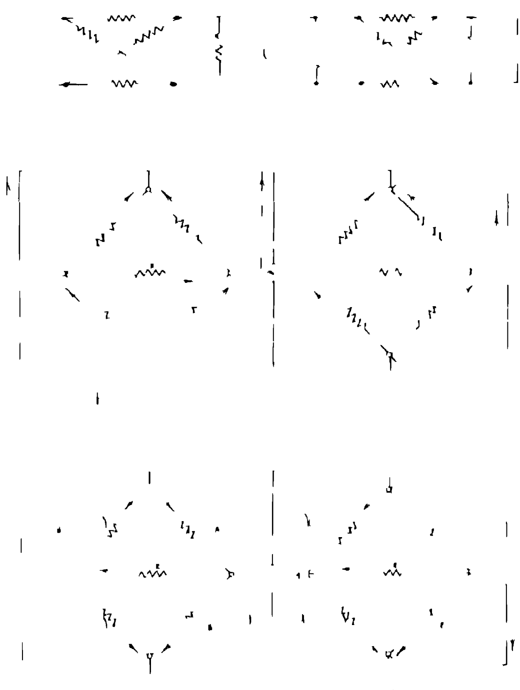


FIG. 607. Development of lattice section with shunt impedances.

Also, from equation 107, putting $E_2 Y_0 = -i$ one obtains

$$E(Y_A - Y_0) = E_2(Y_A + Y_0)$$

$$\gamma = \log_e \frac{E}{E_2}$$

$$= \log_e \frac{Y_A + Y_0}{Y_A - Y_0}$$

$$\gamma = \log_e \frac{Z_A^{-1} + Z_0^{-1}}{Z_A^{-1} - Z_0^{-1}}$$

$$= \log_e \frac{Z_0 + Z_A}{Z_0 - Z_A} \quad \text{(as seen on p. 587)} \quad (109)$$

Modified lattice section (Fig. 607*f*).—

Terminal 1 $\quad i = YE = (Y_B - Y)(I_1 + I_2) + (Y_A - Y)I_1$

i.e. $\quad i = IY + I_1(Y_A + Y_B - 2Y) + I_2(Y_B - Y) \quad (110)$

Terminal 2 $\quad i = YI - (Y_A - Y)(I - I_1 - I_2)$
$\quad + (Y_B - Y)(I - I_1)$

i.e. $\quad i = I(Y_A + Y_B - Y) - I_1(Y_A + Y_B - 2Y)$
$\quad - I_2(Y_A - Y) \quad (111)$

Terminal 3 $\quad 0 = (Y_A - Y)I_1 - (Y_A + Y)I_2$
$\quad - (Y_B - Y)(I - I_1)$

i.e. $\quad 0 = -I(Y_B - Y) + I_1(Y_A + Y_B - 2Y)$
$\quad - I_2(Y_A + Y) \quad (112)$

From (110) and (112):—

$$i = IY_B + I_2(Y_A - Y_B) \quad (113)$$

From (111) and (112)

$$i = I_2Y_A + I(Y_A - Y_B) \quad (114)$$

Equations 113 and 114 are seen to be identical with equations 106 and 107 respectively and they will therefore yield the same results for Z_0 and γ. The modified lattice section of Fig. 607*d* is therefore electrically equivalent to the original lattice section of Fig. 607*a*.

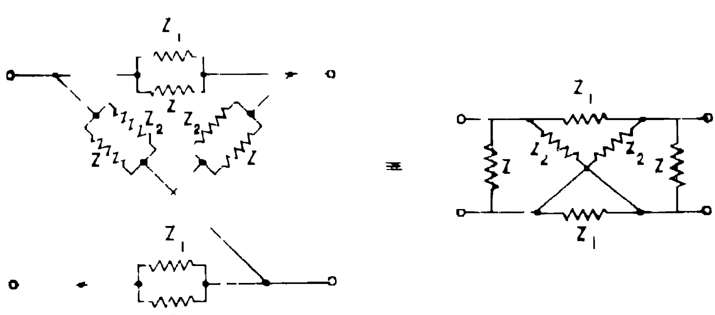


Fig. 608.—Lattice section and its equivalent developed section with shunt impedances.

In practice, this theorem is most useful when all four arms of a lattice section have a common shunt impedance. This is illustrated in Fig. 608, where $Z_A = Y_A^{-1}$ consists of Z_1 and Z in parallel, and $Z_B = Y_B^{-1}$ consists of Z_2 and Z in parallel.

Equivalence of a lattice to a π section

Theorem. A lattice section can be interchanged in any network with a π section and *vice versa* provided that certain relations are maintained between the elements of the two sections.

This theorem follows from the previous theorem. No restriction was imposed on the value of the admittance Y that was subtracted from the series and lattice arms and placed across the input and output terminals. Y can therefore be made equal to Y_B.

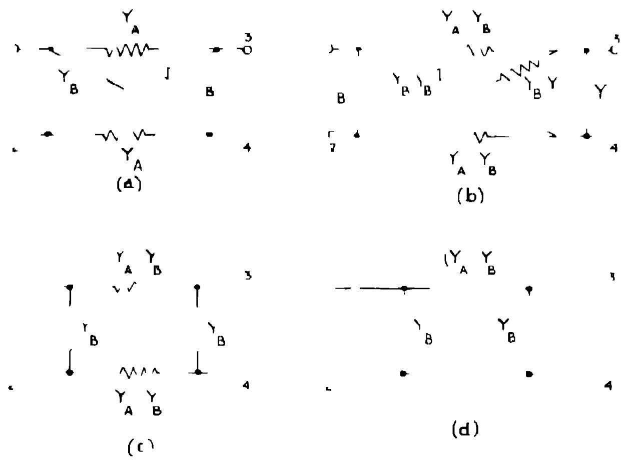


Fig. 609. Equivalence of lattice and π section.

When Y_B is subtracted from both series and lattice arms of the section shown in Fig. 609*c* and placed in shunt with both the input and the output terminals, the admittance of each of the lattice arms is seen to be zero (Fig. 609*c*) and the resulting section can be redrawn as in Fig. 609*c*. This is a balanced π section, the unbalanced form of which is shown in Fig. 609*d*.

The characteristic impedance and propagation constant of the π section shown in Fig. 609*d* are given by

$$Y_{0\pi} = \frac{1}{Z_{0\pi}} = \sqrt{\ldots Y_A \; Y_B\, 2Y_1 \; \tfrac{1}{4}(2Y_1)} = \sqrt{Y_A Y_B} \tag{115}$$

$$Z_0 = \frac{1}{\sqrt{Y_A Y_B}} = \sqrt{Z_A Z_B} \tag{116}$$

and

$$\gamma = \log_e \left(1 + \frac{2Y_1}{Y_A - Y_B} + \frac{2Y_0}{Y_A - Y_B}\right) = \log_e \frac{Y_A + 2Y_0 + Y_B}{Y_A - Y_B}$$

$$\gamma = \log_e \frac{\sqrt{Y_A} + \sqrt{Y_B}}{\sqrt{Y_A} - \sqrt{Y_B}} = \log_e \frac{Y_A + Y_0}{Y_A - Y_0} \tag{117}$$

$$= \log_e \frac{Z_A^{-1} + Z_0^{-1}}{Z_A^{-1} - Z_0^{-1}} = \log_e \frac{Z_0 + Z_A}{Z_0 - Z_A} \tag{118}$$

These results are identical with those obtained on page 587 for the original lattice section of Fig. 609*a* and therefore the two sections are electrically equivalent.

CHAPTER 14

ATTENUATION AND ATTENUATORS

EXPRESSION OF ATTENUATION IN DECIBELS AND IN NEPERS

The *decibel* is fundamentally a unit of power ratio, but as has been shown in Chapter 5, it can be used to express current ratios when the resistive components of the impedances through which the current flows are equal, and voltage ratios when the conductive components of these impedances are equal. The *neper* is fundamentally a unit of current ratio, but it can be used to express power ratios when the resistive components of the impedances are equal.

The loss of power in a transmission line or electrical network is known as attenuation. Attenuation may be measured using either the decibel or the neper notation.

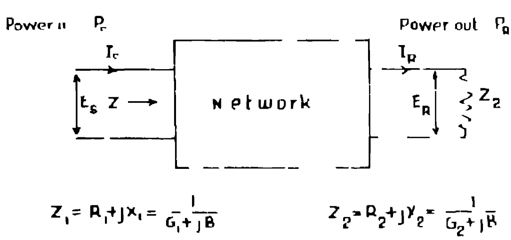


Fig. 610. Attenuation measurement in decibel and neper notations

If the power entering a network is P_S and the power leaving it is P_R (see Fig. 610), then the attenuation in decibels is defined as

$$\text{Attenuation in decibels} = 10 \log_{10} \left| \frac{P_S}{P_R} \right| \tag{1}$$

If the current entering a network is I_S and the current leaving it is I_R, then the attenuation in nepers is defined as

$$\text{Attenuation in nepers} = \log_e \left| \frac{I_S}{I_R} \right| \tag{2}$$

Because of its derivation from the exponential e, the neper is the most convenient unit for expressing attenuation in theoretical work. The decibel on the other hand being defined in terms of logarithms to base 10 is a more convenient unit in practical calculations using the decimal system of reckoning. The conditions under which the two units may be used can be summarised in the

following equations, the notation of which is indicated in Fig. 610

$$\text{Attenuation in db} = 10 \cdot \log_{10}\left|\frac{P_S}{P_R}\right| \qquad (3)$$

$$= 20 \cdot \log_{10}\left|\frac{I_S}{I_R}\right| \text{ (provided that } R_1 = R_2\text{)} \qquad (4)$$

$$= 20 \cdot \log_{10}\left|\frac{E_S}{E_R}\right| \text{ (provided that } G_1 = G_2\text{)} \qquad (5)$$

$$\text{Attenuation in nepers} = \log_e\left|\frac{I_S}{I_R}\right| \qquad (6)$$

$$= \log_e\left|\frac{E_S}{E_R}\right| \text{ (provided that } |Z_1| = |Z_2|\text{)} \qquad (7)$$

$$= \tfrac{1}{2}\log_e\left|\frac{P_S}{P_R}\right| \text{ (provided that } R_1 = R_2\text{)} \qquad (8)$$

If the resistive components of the impedances at the input and output of the network are equal, then the attenuation may be readily converted from one notation to the other, for

$$(\text{Attenuation in db}) = 20 \cdot \log_{10}\left|\frac{I_S}{I_R}\right|$$

$$= 20 \cdot \log_e\left|\frac{I_S}{I_R}\right| \times \log_{10} e$$

$$= 8 \cdot 686 \log_e\left|\frac{I_S}{I_R}\right|$$

$$= 8 \cdot 686 \times (\text{attenuation in nepers})$$

Thus :—

$$\text{Attenuation in db} = 8 \cdot 686 \times \text{attenuation in nepers (provided that } R_1 = R_2\text{)} \qquad (9)$$

$$\text{Attenuation in nepers} = 0 \cdot 1151 \times \text{attenuation in db (provided that } R_1 = R_2\text{)} \qquad (10)$$

ATTENUATING NETWORKS

In transmission equipment, it is frequently desired to attenuate the currents and voltages at certain stages. Attenuators and pads are networks designed to meet this requirement, and since to prevent attenuation distortion, all frequencies must be attenuated to the same degree the networks must consist of purely resistive components. No phase-shift will be introduced by such networks; thus, for each network, the phase constant (β) will be zero, and the propagation constant (γ) will simply be equal to the attenuation constant (α).* A fixed attenuator is sometimes known as a "pad".

* Bearing these facts in mind, the designs and properties of attenuating networks may be deduced from the equations of Chapter 13. In this chapter, however, the results will, in many cases, be obtained in a simpler manner from first principles for the benefit of the readers less familiar with the subject.

These networks, by choice of suitable resistances, may have any required value of attenuation. They may be designed to have any resistive value of characteristic impedance, if symmetrical, or of image impedances if asymmetrical. One of these networks may therefore be used in place of a transformer for matching between circuits of different resistive impedance, thus avoiding, particularly in carrier-frequency circuits, the attenuation distortion introduced by a transformer. The attenuation introduced will be of little consequence if valve amplification is included in the circuit.

There are three conditions that the attenuating network must fulfil. It must give :—

(1) the correct input impedance ;
(2) the correct output impedance ;
(3) the specified attenuation.

This attenuation is usually quoted in decibels :-

Attenuation in decibels $\qquad D = 10 \log_{10} \frac{P_S}{P_R}$

where P_S is power input and P_R power output

In the following considerations the symbol N will be used for $\sqrt{\frac{P_S}{P_R}}$, *i.e.* : –

$$\text{Attenuation } D = 20 \log_{10} \sqrt{\frac{P_S}{P_R}} = 20 \log_{10} N \qquad (11)$$

If both pairs of terminals of the network are matched to the same impedance, then :—

$$\frac{P_S}{P_R} = \frac{I_S^2}{I_R^2}$$

Therefore for a pad in a symmetrical circuit, $N = \frac{I_S}{I_R}$, but in an asymmetrical circuit the value $N = \sqrt{\frac{P_S}{P_R}}$ must always be used.

Symmetrical T type

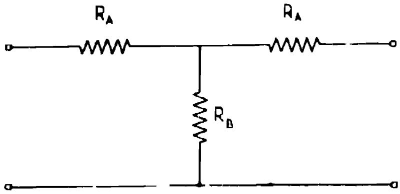


FIG. 611.—Symmetrical T network.

This is one of the most common types of pad, and consists of a divided series arm and one central shunt arm. The pad used between equal impedances will be symmetrical, *i.e.*, the series arm is divided into two equal parts (*see* Fig. 611). The values of the

series and shunt arms for a given value of impedance and attenuation will now be determined

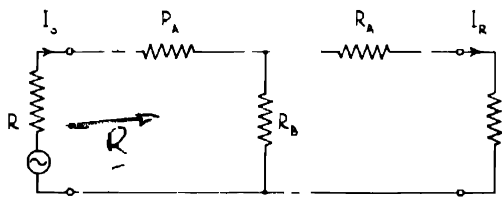


FIG 612 —Symmetrical T network

Consider the input current I_S (Fig 612) At the shunt arm it divides in proportion to the conductances

Hence $$I_R = \frac{R_B}{R_B + R_A + R} I_S \tag{12}$$

$$\therefore \quad N = \frac{I_S}{I_R} = \frac{R_B + R_A + R}{R_B} \tag{13}$$

But the impedance looking into the attenuator is required to be R

Hence $$R = R_A + \frac{R_B(R_A + R)}{R_B + R_A + R}$$

$$= R_A + \frac{R_A + R}{N}$$

$$\therefore \quad R(N - 1) = R_A(N + 1)$$

$$\therefore \quad R_A = R\left(\frac{N-1}{N+1}\right) \tag{14}$$

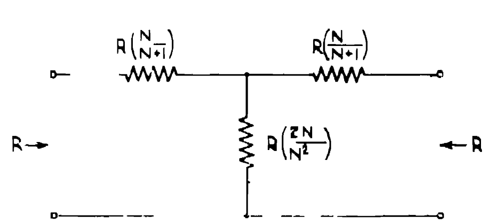


FIG 613 —T network having input and output impedance equal to R

But $$N = \frac{R_B + R_A + R}{R_B}$$

$$\therefore \quad R_B(N - 1) = R_A + R$$

$$= R\left(\frac{2N}{N+1}\right)$$

$$\therefore \quad R_B = R\left(\frac{2N}{N^2 - 1}\right) \tag{15}$$

Using these formulae, therefore, an attenuator can be designed to give the specified attenuation, and to be properly matched to the circuit

The resultant T section is shown in Fig 613.

It may be noted that the results obtained also follow directly from Fig 574 (Chapter 13, page 575), bearing in mind that in this case the characteristic impedance of the required T section is R and the propagation constant is α

$$\text{i.e.}\quad R_A = R\tanh\frac{\alpha}{2} = R\frac{e^{\frac{\alpha}{2}} - e^{-\frac{\alpha}{2}}}{e^{\frac{\alpha}{2}} + e^{-\frac{\alpha}{2}}} = R\frac{e^{\alpha} - 1}{e^{\alpha} + 1} = R\left(\frac{N-1}{N+1}\right)$$

$$\text{and}\quad R_B = \frac{R}{\sinh\alpha} = \frac{2R}{e^{\alpha} - e^{-\alpha}} = \frac{2Re^{\alpha}}{e^{2\alpha} - 1} = R\left(\frac{2N}{N^2 - 1}\right)$$

$$\text{where}\quad e^{\alpha} = \frac{I_S}{I_R} = N$$

Example Design a T type pad to give 25 db attenuation and to have a characteristic impedance of 600 ohms

$$N = \text{antilog}_{10}\frac{D}{20} = \text{antilog}_{10}\frac{25}{20}$$
$$= 17{\cdot}8$$
$$R_A = R\left(\frac{N-1}{N+1}\right)$$
$$= 600 \times \frac{16{\cdot}8}{18{\cdot}8}$$
$$= 536 \text{ ohms}$$
$$R_B = 2R\left(\frac{N}{N^2-1}\right)$$
$$= 1200 \times \frac{17{\cdot}8}{316}$$
$$= 67{\cdot}6 \text{ ohms.}\quad \textit{Ans.}$$

Asymmetrical T type

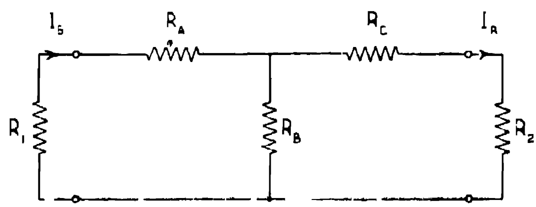


FIG 614—Asymmetrical T section.

Fig. 614 shows a pad that is not symmetrical.

Here $$N = \sqrt{\frac{P_S}{P_R}} = \sqrt{\frac{I_S^2 R_1}{I_R^2 R_2}} \qquad (16)$$

Using this and formulae for the input and output impedance it can be shown that —

$$R_A = R_1\left(\frac{N^2+1}{N^2-1}\right) - 2\sqrt{R_1R_2}\left(\frac{N}{N^2-1}\right) \qquad (17)$$

$$R_B = 2\sqrt{R_1R_2}\left(\frac{N}{N^2-1}\right) \qquad (18)$$

$$R_C = R_2\left(\frac{N^2+1}{N^2-1}\right) - 2\sqrt{R_1R_2}\left(\frac{N}{N^2-1}\right) \qquad (19)$$

A network consisting of these components will have image impedances R_1 and R_2. If $R_1 > R_2$ then it will be found that $R_A > R_C$.

L type

If a pad is required for matching purposes only, then the design will be such as to give minimum attenuation. Examining the T section it will be seen that this condition will be reached when R_C has been reduced to zero.

This then forms the L type pad (see Fig. 615).

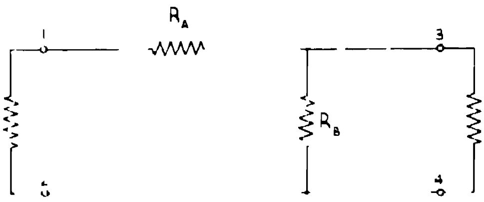


Fig. 615.—L type network.

To obtain values for R_A and R_B, consider input impedances. Looking in at terminals 1 and 2 —

$$R_1 = R_A + \frac{R_2R_B}{R_2 + R_B}$$

$$\therefore\ R_1R_B + R_1R_2 = R_AR_B + R_2R_A + R_2R_B \qquad (20)$$

and looking in at terminals 3 and 4 —

$$R_2 = \frac{R_B(R_1 + R_A)}{R_B + R_1 + R_A}$$

$$R_2R_B + R_2R_A + R_1R_2 = R_BR_A + R_1R_B \qquad (21)$$

Adding equations 20 and 21 : —

$$2R_1R_2 = 2R_AR_B$$

or $$R_A = \frac{R_1R_2}{R_B}$$

Substituting in (20) —

$$R_B = \sqrt{\frac{R_1 R_2^2}{R_1 - R_2}} \tag{22}$$

Hence $$R_A = \sqrt{R_1(R_1 - R_2)} \tag{23}$$

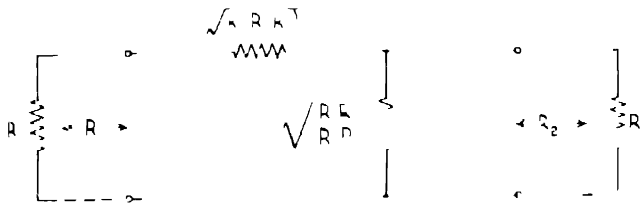


FIG 616 —L network having image impedances R_1 and R_2

Fig. 616 shows the resultant L type network. It is a network having image impedances R_1 and R_2

π type

This attenuator is another common type, consisting of one series arm and two shunt arms. When used purely as an attenuator between equal impedances, symmetry demands that these two shunt arms shall be equal

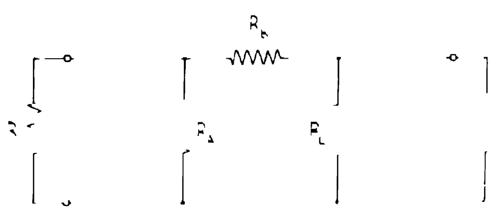


FIG. 617 —π networks

By calculation similar to that used for the L section it may be shown that

$$R_A = R_C = R\left(\frac{N+1}{N-1}\right) \tag{24}$$

$$R_B = R\left(\frac{N^2-1}{2N}\right) \tag{25}$$

In a pad used for matching $R_A \neq R_C$ and the following formulae apply —

$$R_A = R_1\left(\frac{N^2-1}{N^2-2NS+1}\right) \tag{26}$$

$$R_B = \frac{\sqrt{R_1R_2}}{2}\left(\frac{N^2-1}{N}\right) \tag{27}$$

$$R_C = R_2\left(\frac{N^2 - 1}{N^2 - 2\frac{N}{S} + 1}\right)$$

where $S^2 = \frac{R_1}{R_2}$.

There is no difference in the performance of the T and π type pads and each one will suit any requirement, but one will probably be found to have more suitable or standard components than the other. It may be noted that a deviation of 5 per cent. from the calculated values of the resistances will mismatch the impedances by no more than the same amount, and vary the attenuation by as little as 0·5 db.

Balanced T, L and π types

When it is required to balance the two legs of the circuit, as is frequently the case in transmission equipment, then the preceding pads must be modified by dividing the series arm into two equal halves and inserting one half in each leg (*see* Fig. 618).

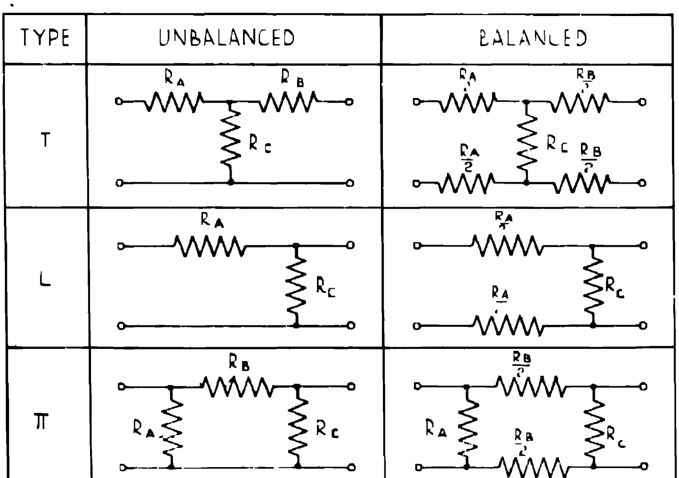


Fig. 618.—Balanced and unbalanced T, L and π networks

When designing these balanced pads the components of the unbalanced type should be calculated using the formulae already quoted, and the series arm divided between the two legs. The characteristics of this derived pad—that is, the impedance and the attenuation—will be identical to those of the unbalanced pad.

Bridged-T type

Fig. 619 shows a symmetrical bridged-T type section used between equal impedances.

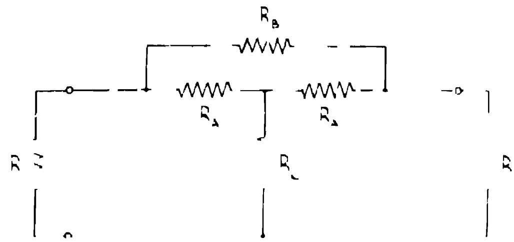


FIG. 619.—Symmetrical bridged-T network.

The network may be designed to have a constant impedance R, but any desired attenuation by making :—

$$R_B R_C = R_A^2 = R^2$$

Thus to vary the attenuation without changing the design impedance, only two resistances have to be varied *viz.* R_B and R_C. It should be noted that, in the case of a symmetrical T or π section attenuator, three resistances have to be varied to change the attenuation without altering the impedance.

The design formulae for the bridged-T section are :—

$$R_A = R \qquad (29)$$

$$R_B = R(N - 1) \qquad (30)$$

$$R_C = \frac{R_A^2}{R_B} = \frac{R}{N - 1} \qquad (31)$$

Example.

Design a bridged-T attenuator having an attenuation of 40 db when working between two 600 ohms impedances.

To give 40 db attenuation, $N = 100$

$R_A = 600$ ohms *Ans.*

$R_B = 600\ (100 - 1)$

$= 59\ 400$ ohms *Ans.*

$R_C = \dfrac{600}{(100 - 1)}$

$= 6 \cdot 06$ ohms *Ans.*

The network is shown in Fig. 620.

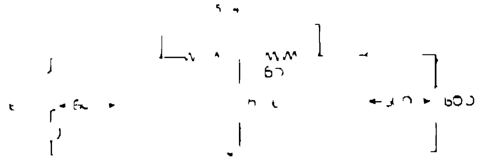


FIG. 620.—Bridged-T network having an attenuation of 40 db.

Lattice type

This type, shown in Fig. 621, is occasionally used.

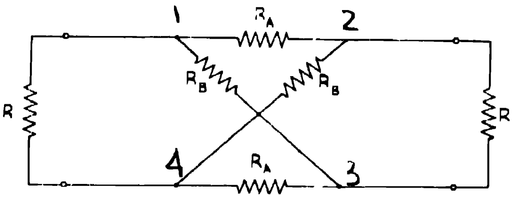


FIG. 621.—Lattice network.

The characteristic impedance R of the network can best be determined by consideration of its open-circuit and short-circuit impedances.

$$R_{oc} = \frac{R_A + R_B}{2} \tag{32}$$

$$R_{sc} = 2\frac{R_A R_B}{R_A + R_B} \tag{33}$$

$$\therefore \quad R = \sqrt{R_{oc} R_{sc}}$$
$$= \sqrt{R_A R_B} \tag{34}$$

This gives the condition for matching the network to its adjacent circuit. It can, of course, be used only in a symmetrical case.

It may be shown that—

$$N = \frac{I_S}{I_R} = \frac{R_A + R_B + 2R}{R_B - R_A}$$

But, from equation 34, $R_B = \dfrac{R^2}{R_A}$

$$\therefore \quad N = \frac{R_A{}^2 + R^2 + 2R\,R_A}{R^2 - R_A{}^2}$$
$$= \frac{R + R_A}{R - R_A} \tag{35}$$

Hence

$$R_A = R\left(\frac{N - 1}{N + 1}\right) \tag{36}$$

$$R_B = R\left(\frac{N + 1}{N - 1}\right) \tag{37}$$

DESIGN OF ATTENUATORS AND PADS

The steps in the design of any pad may be summarised as follows:—

(1) Determine the type of pad to be used. In some cases alternative networks will be possible, and that one should be chosen which gives the most convenient component values.

(2) Change the required decibel attenuation D to ratio N by use of Table XVII or the graph in Fig 622

(3) If a T or π network is to be used for matching between unequal impedances R_1 and R_2 verify that the value of attenuation chosen is greater than the minimum permissible attenuation given by Fig 623 otherwise one arm of the network will work out to be negative With the L type pad this is unnecessary

(4) With the information so obtained and a knowledge of the design impedance, evaluate the component values by use of the equations already stated above noting that when it is necessary to insert a loss of more than 40 db, it is usually more convenient to use two smaller pads in series.

TABLE XVII

N Values $\left(N = \sqrt{\frac{P}{P_r}}\right)$

D (db)	N	D (db)	N	D (db)	N	D (db)	N
1.0	1.122	18.0	7.943	35.0	56.234	52.0	398.11
2.0	1.259	19.0	8.913	36.0	63.096	53.0	446.68
3.0	1.412	20.0	10.000	37.0	70.795	54.0	501.19
4.0	1.585	21.0	11.220	38.0	79.433	55.0	562.34
5.0	1.778	22.0	12.590	39.0	89.125	56.0	630.96
6.0	1.995	23.0	14.125	40.0	100.000	57.0	707.95
7.0	2.239	24.0	15.849	41.0	112.20	58.0	794.33
8.0	2.512	25.0	17.783	42.0	125.89	59.0	891.25
9.0	2.818	26.0	19.953	43.0	141.25	60.0	1000.0
10.0	3.162	27.0	22.387	44.0	158.49	65.0	1778.3
11.0	3.548	28.0	25.119	45.0	177.83	70.0	3162.3
12.0	3.981	29.0	28.184	46.0	199.53	75.0	5623.4
13.0	4.467	30.0	31.623	47.0	223.87	80.0	10000.0
14.0	5.012	31.0	35.481	48.0	251.19	85.0	17783
15.0	5.623	32.0	39.811	49.0	281.84	90.0	31623
16.0	6.310	33.0	44.668	50.0	316.23	95.0	56234
17.0	7.079	34.0	50.119	51.0	354.81	100.0	10^5

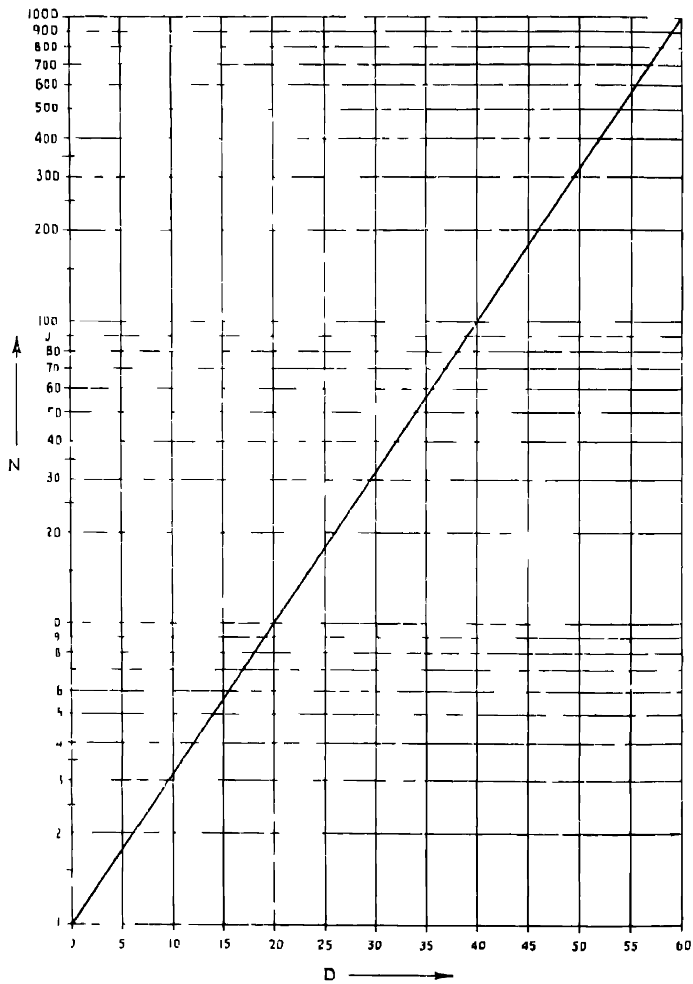


FIG 622 —Graph for converting given decibel loss D to ratio N

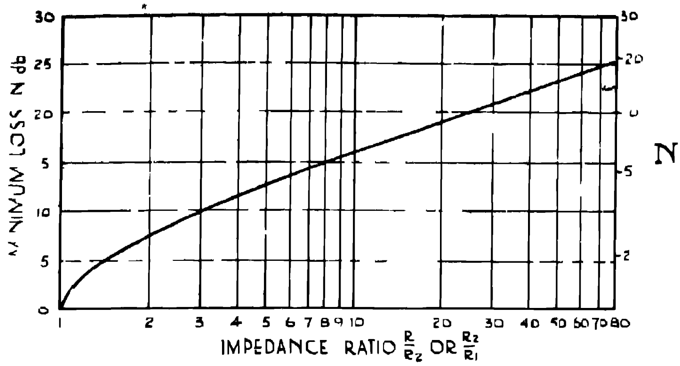


FIG. 623.—Minimum loss in Γ or π networks for given ratios of input and output impedances.

Example.—

Design an attenuating network to match between 400 and 800 ohms, and to give an attenuation of 15 db.

Following the steps indicated, the pad used would be either a Γ or π type. From Fig. 623, $\frac{R_2}{R_1} = 2$ allows a minimum of 7 db attenuation. From the table on p. 615, for 15 db attenuation $N = 5.623$.

(*a*) T type (*see* page 609):—

$$R_A = R_1\left(\frac{N^2+1}{N^2-1}\right) - 2\sqrt{R_1R_2}\left(\frac{N}{N^2-1}\right)$$

$= 218$ ohms. *Ans.*

$$R_B = 2\sqrt{R_1R_2}\left(\frac{N}{N^2-1}\right)$$

$= 208$ ohms. *Ans.*

$$R_C = R_2\left(\frac{N^2+1}{N^2-1}\right) - 2\sqrt{R_1R_2}\left(\frac{N}{N^2-1}\right)$$

$= 644$ ohms. *Ans.*

The complete T section is illustrated in Fig. 624*b*.

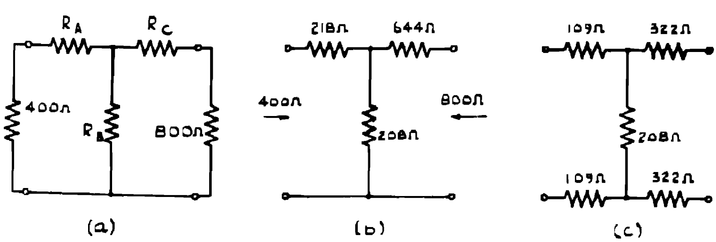


FIG. 624.—T network having input impedance of 400Ω, and output impedance of 800Ω.

The values are quite suitable from a practical viewpoint, ... the pad could now be constructed, dividing R_A and R_C betw... the two arms, as in Fig 624c, if a balanced pad is required.

(b) π type (see page 611)

$$R_A = R_1\left(\frac{N^2 - 1}{N^2 - 2N S + 1}\right)$$

497 ohms *Ans*

$$R_B = \frac{\sqrt{R_1 R_2}}{2}\left(\frac{N^2 - 1}{N}\right)$$

1540 ohms *Ans*

$$R_C = R_2\left(\frac{N^2 - 1}{N^2 - 2\frac{N}{S} + 1}\right)$$

= 1463 ohms *Ans*

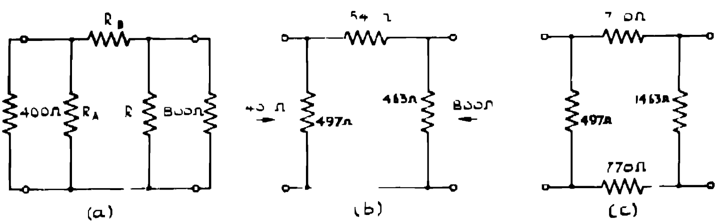


Fig 625 π network having input impedance of 400Ω and output impedance of 800Ω

The complete π section is shown in Fig 625b Fig 625c shows the corresponding balanced form It will be noted that there is a considerable difference in the resistance values for the T and π types and the more convenient may be used

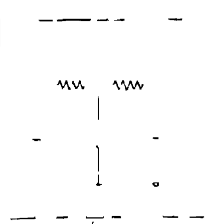
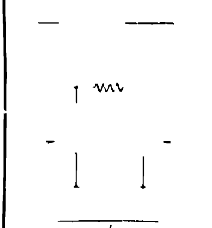
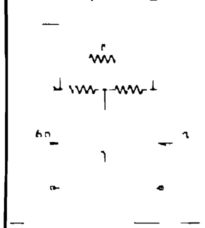
Components for attenuators in 600 ohm circuits

As 600 ohms is the characteristic impedance of most line communication circuits, the majority of attenuators and pads will come under this heading Figs 626 and 627 give graphs showing component values for T and π networks respectively for use in such circuits Table XVIII gives the component values for T, π and bridged T networks

If the characteristic impedance is not 600 ohms, but R, the values for the components must all be multiplied by $\frac{R}{600}$.

TABLE XVIII

Pads designed for 600 ohms characteristic impedance

Loss D in db	T pad		π pad		Bridged T pad	
						
	a	b	c	d	e	f
1	34.50	5201	69.2	10 436	73.2	4918
2	68.79	2583	139.4	5233	155	2317
3	102.5	1703	211.1	3512	247	1456
4	135.8	1258	286.2	2651	351	1025
5	168.0	987.1	365.0	2142	467	771
6	199.4	803.4	448.1	1806	597	603
7	229.4	670.0	537.3	1569	743	485
8	258.3	567.5	631.1	1394	907	397
9	285.7	487.1	738.9	1260	1091	330
10	311.7	421.9	853.1	1155	1297	278
11	336.3	367.2	980.3	1071	1530	235
12	359.1	321.7	1119	1003	1789	201
13	380.5	282.7	1273	946.1	2080	173
14	400.4	249.3	1444	899.1	2407	149
15	418.8	220.1	1633	859.5	2773	130
20	490.9	121.2	2970	733.3	5400	66.7
25	536.1	67.61	5324	671.4	10 070	35.8
30	563.2	37.99	9486	639.0	18 370	19.6
35	579.0	21.35	16 864	621.6	33,140	10.9
40	589.1	12.00	30,000	612.1	59 400	6.06
45	593.3	6.748	53,350	606.8	106,100	3.40
50	596.2	3.795	94,860	603.8	189,100	1·90

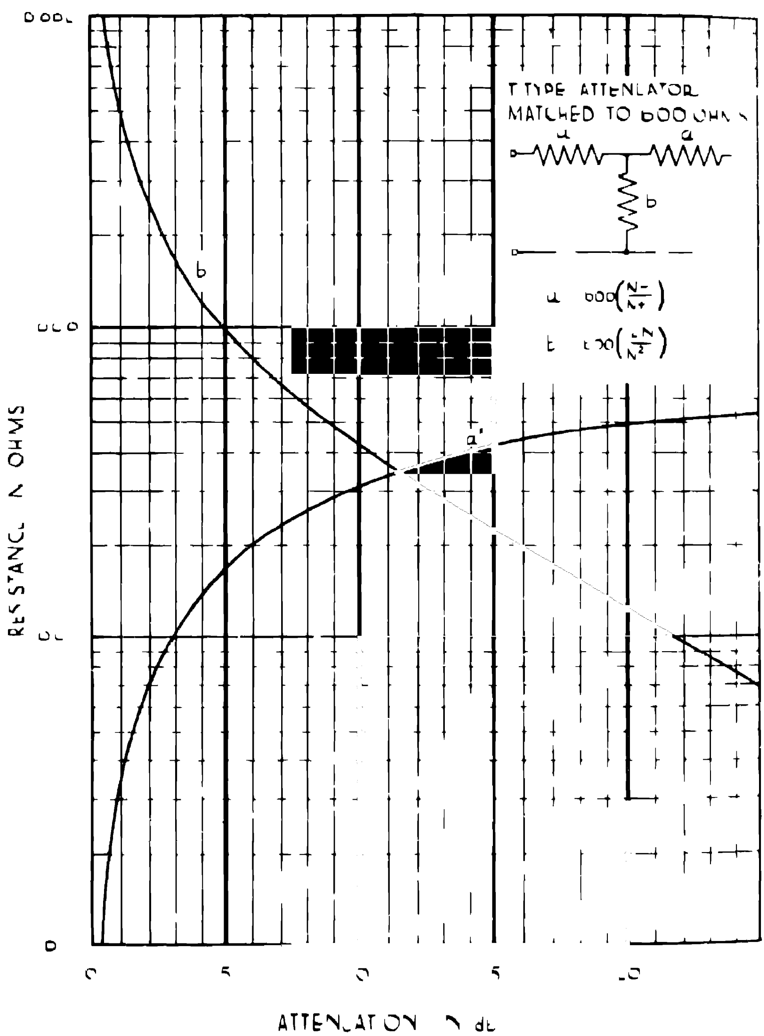


FIG 626 Graph giving component values for T network (characteristic impedance 600 Ω)

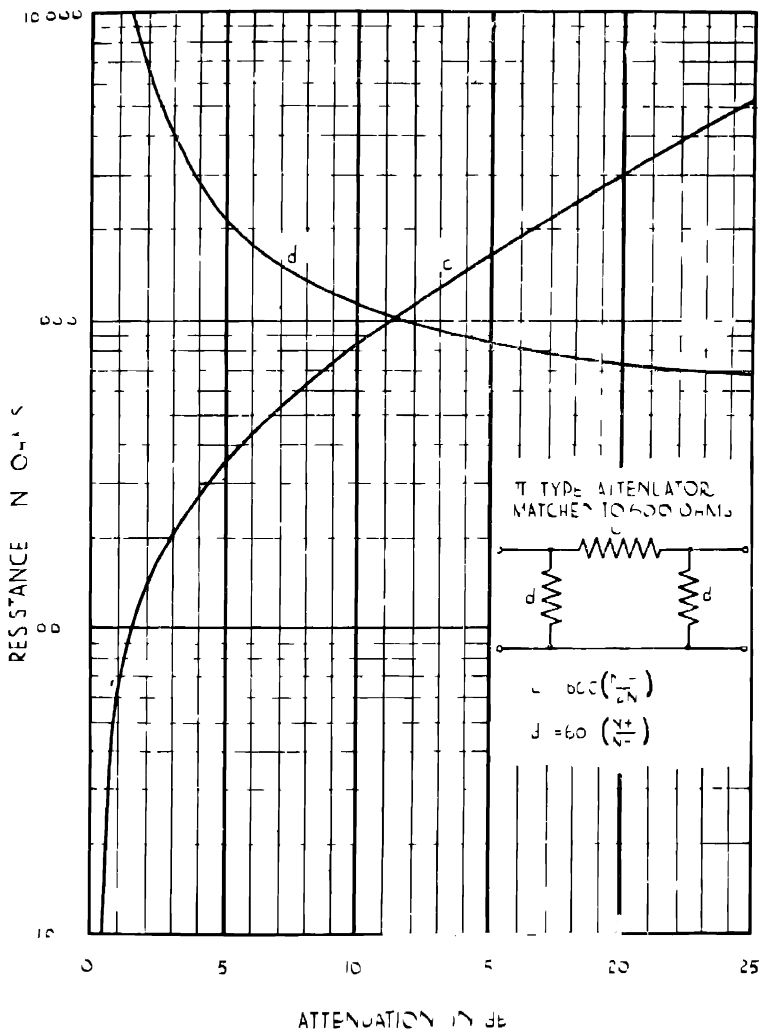


FIG 627 —Graph giving component values for π network (characteristic impedance 600 Ω)

VARIABLE ATTENUATORS

Variable attenuators are so designed as to have a const input and output impedance, but a variable attenuation Th may be divided into several classes depending on the method achieving the result

The elementary type has the simple construction of a T π section and the resistors are variable All are ganged togeth so that at different positions the pad impedance is unalter although the attenuation is varied (*see* Fig 628*a* and *b*)

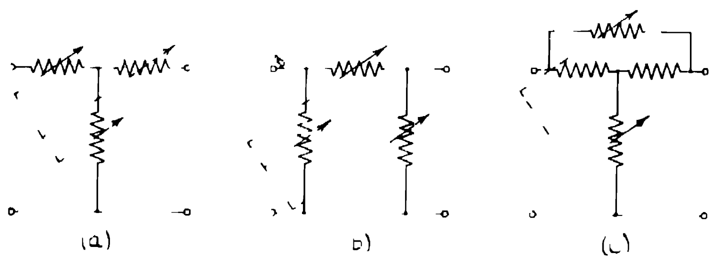


FIG 628 —Variable attenuator

The bridged-T type (*see* Fig 628*c*) has already been discussed This has the advantage compared with those mentioned above that only two resistors have to be varied, as compared with three when T or π sections are used

A further type, simple in construction and design consists of a number of pads of equal impedance but different attenuation connected in series Each pad may be switched in or out as required

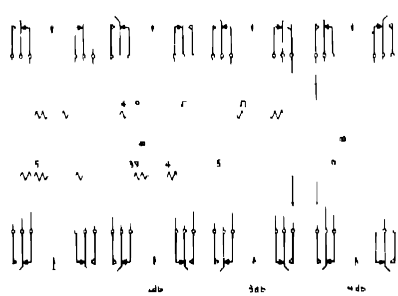


FIG 629 Non reactive adjustable attenuator (characteristic impedance 600 Ω)

Fig 629 shows such an attenuator

Each pad is of the balanced T type and may be brought into circuit by operation of the appropriate key. The resistances are

shown in ohms, and the characteristic impedance of each pad is 600 ohms.

Fig. 630 shows how the principle may be extended so that, with appropriate switching, only three pads, namely 5 10 and 20 db, need be employed to form a variable attenuator covering the range from 0 to 30 db in 5 db steps.

S1

Posn.	Att. db.
1	0
2	5
3	10
4	15
5	20
6	25
7	30

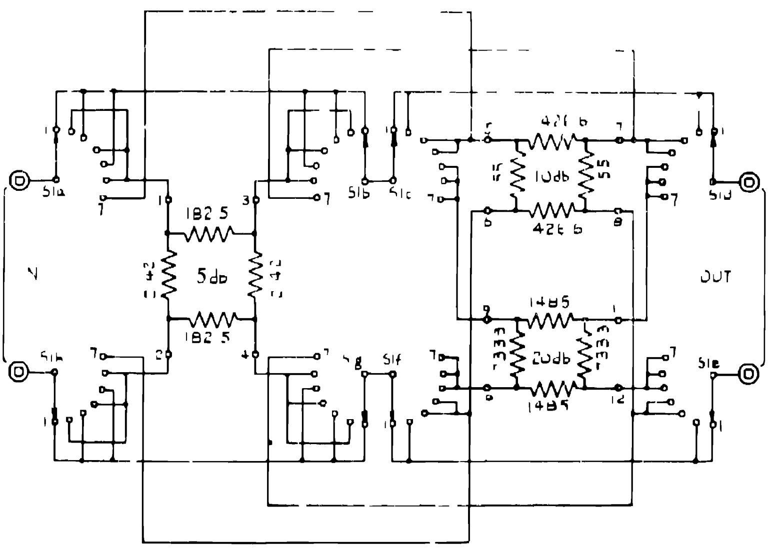


FIG. 630.—Non-reactive adjustable attenuator : Range 0–30 db.

TRANSMISSION MEASUREMENTS

It has been seen (Chap. 5) that the power level in a circuit may be expressed using the decibel notation provided that a reference power is stated ; in line communication this reference power is taken as 1mW. A decibel-meter, which is the name given to an instrument measuring power levels when the decibel notation is used, should be a specially calibrated wattmeter. However, it is found impossible in practice to design a wattmeter that is sufficiently accurate at all frequencies over the required range. On the other hand if all measurements are taken at points having the same standard impedance, the voltage will give a direct

indication of the power and a suitable voltmeter may be recalibrated to read the power level directly in decibels. This is the basis of all transmission measurements.

The standard impedance usually chosen is a pure resistance of 600 ohms. All testing apparatus is therefore designed to work into this 600 ohms impedance. All points at which tests have to be made must be designed to have this impedance or else a correction factor will have to be applied.

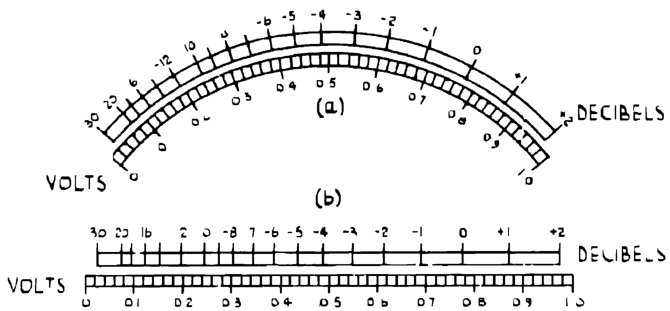


FIG. 631 —Calibration of voltmeter to read power levels above and below 1mW in 600Ω.

Since the reference power is 1mW, and this is applied to a 600 ohms resistance, the reference voltage level will be :—

$$\sqrt{0\cdot001 \times 600} = 0\cdot775 \text{ volts.}$$

This corresponds to a current in 600 ohms of 1·29mA.

If E is a voltage reading, the corresponding power level in decibels, reference 1mW and 600Ω, will be given by :—

$$\text{Power level} = 20 \log_{10} \frac{E}{0\cdot775}$$

Fig. 631 shows the corresponding voltage and dbm readings.

Measurement of power levels

The reading of power level at a point in a circuit may be obtained in two ways :—

(*a*) *" Level " or " Through " measurement.*

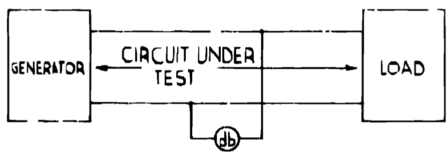


FIG. 632.—Decibelmeter giving " Level " or " Through " measurement.

A "level" measurement is obtained by tapping a high-impedance voltmeter across the circuit, as in Fig 632. The high impedance of the meter is essential to ensure that its presence will not disturb the circuit under test (*e g.*, a 5000 ohm meter will introduce a shunt loss of about 0·5 db). If the impedance of the circuit under test is 600 ohms, the voltmeter, calibrated to read power level directly in decibels, will give a true reading Any variation in circuit impedance from 600 ohms will of course, destroy the accuracy of the measurement, but, provided that the impedance is known, a correction factor may be applied.

Let the voltage be E and let the circuit impedance be 400 ohms instead of 600 ohms

$$\text{Power indicated on scale} = \frac{E^2}{600}$$

$$\text{Actual power} = \frac{E^2}{400}$$

$$\text{Error} = 10 \log_{10} \frac{400}{600}$$

$$= 1\cdot 76 \text{ db}$$

i e, the meter will read 1·76 db low.

(*b*) "*Transmission*" "*Terminating*" or "*Loss*" *measurement*

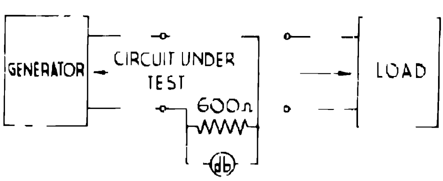


FIG 633 —Decibelmeter giving " Transmission " measurement

A "transmission" (TRANS) measurement is made by terminating the circuit in a 600 ohm resistance, and measuring the voltage across it using the meter, Fig 633.

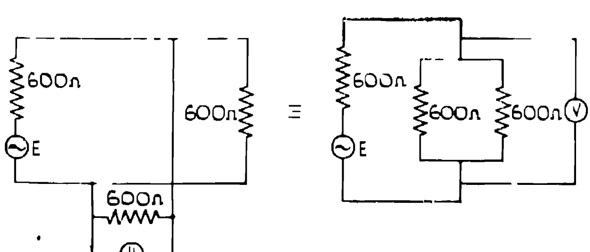


FIG. 634.—"Level" measurement with meter set at " TRANS ".

Many decibel-meters give both facilities, a high impedance meter being used, the 600 ohms resistance for TRANS measurements being brought into circuit by operation of a switch. Failure to cut out the 600 ohms when making a LEVEL measurement will give a reading that is 3·52 db too low (*See* Fig. 634.)

Whereas voltage reading should be $\frac{E}{2}$, it will be $\frac{E}{3}$;

$$\therefore \text{ discrepancy of reading } = 20 \log_{10} \frac{2}{3}$$
$$= -3{\cdot}52 \text{ db.}$$

Decibel-meters

Especially when making LEVEL measurements it is essential that the decibel-meter used shall take only the minimum possible power from the circuit under test. Since such a meter is expected to give a reading when the power in the circuit under test is of the order of, say, 0·1mW, it follows that an extremely sensitive meter movement must be used, unless some form of valve amplification is to be employed. Such a sensitive meter has the disadvantage that its movement may be easily damaged by accidental over loading. Decibel-meters may be divided into two classes according to whether or not they utilise valve amplification.

Metal rectifier type decibel-meter.—This type of decibel-meter is frequently used owing to its simplicity. In most cases a moving coil meter movement is used. The incoming signal is rectified by metal rectifiers in the form of a full-wave rectifier bridge circuit, and the rectified current is passed through the meter.

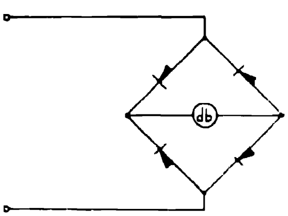
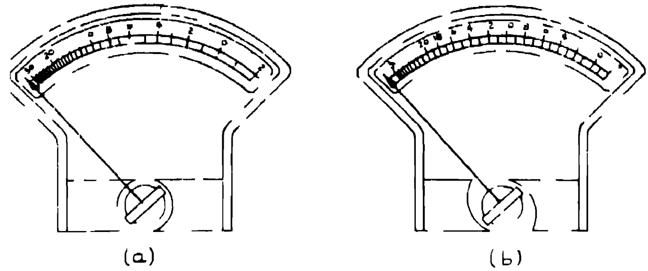


FIG. 635.—Arrangement of rectifiers in db meter

In order to spread out the scale at the lower part of the range specially shaped pole-pieces may be employed in the meter as shown in Fig. 636.

Fig. 637 shows a decibel-meter suitable for use over the audio range (say up to about 5000 c/s). The frequency range is limited in this case by the input transformer. The meter will give TRANS or LEVEL readings, and the scale is calibrated reference 1mW in 600Ω to read from − 15 to 0 dbm. To enable higher power



(a) Uncorrected (b) Corrected

FIG. 636.—Illustrating the effect of shaped pole pieces in db meter

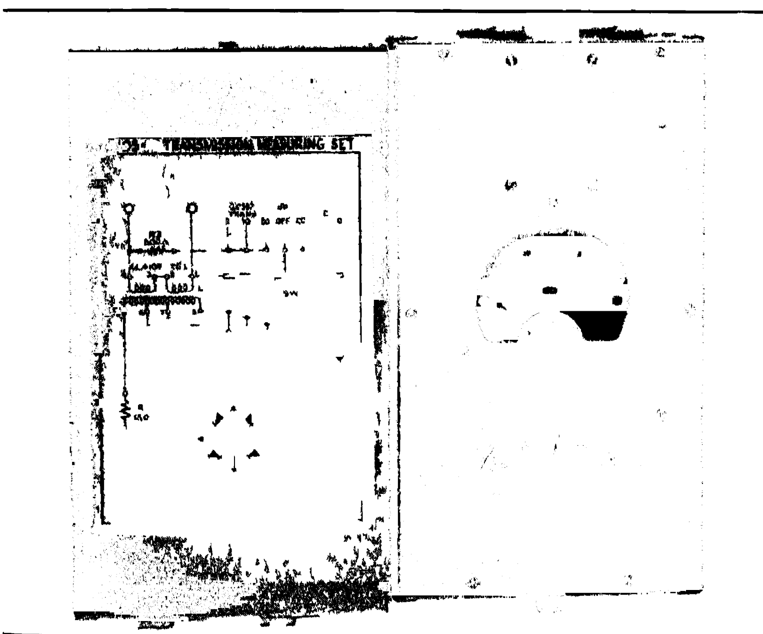


PLATE 31.—Metal-rectifier type decibel-meter (TMS No. 2)

levels to be measured, tappings are provided on the input transformer introducing attenuations of 10 or 20 db into the input circuit to the meter thus providing two new ranges — 5 to + 10 dbm and + 5 to + 20 dbm

The accuracy of such a meter depends on the frequency error of the rectifiers used but by using rectifiers having an extremely small self capacity and omitting the input transformer it is possible to construct instruments having negligible frequency errors up to 50 kc/s

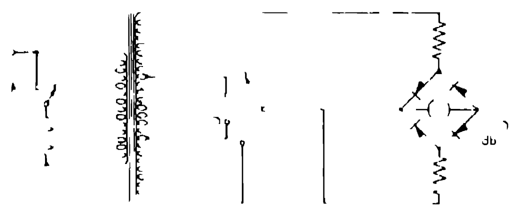


Fig. 637 Decibel meter for use over audio range

The meter movement rectifier network, and associated shunt and series resistors, may be arranged so as to present a 600 ohm impedance at the terminals in which case the meter may be used only for TRANS measurements Alternatively, the circuit may be arranged so as to offer a very high shunt impedance across the circuit under test allowing LEVEL measurement to be made

Valve type decibel meters The valve type decibel-meter has two advantages over the type just discussed Firstly, it enables much lower powers to be measured (down to — 40 dbm or lower), and secondly, the circuit may be designed so that the meter movement is not damaged by applying too large a signal at the input.

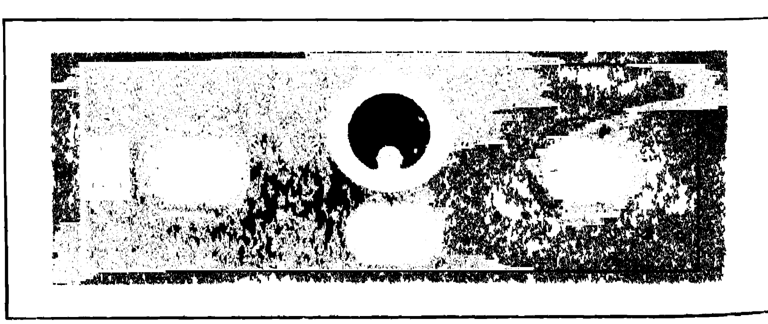


Plate 32 —Valve type decibel-meter

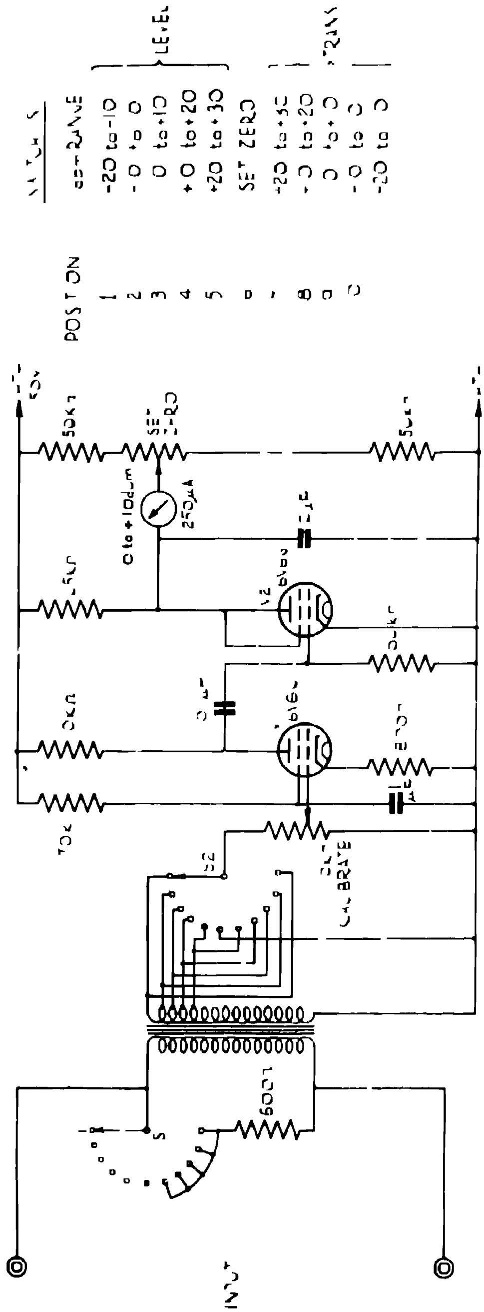


Fig 638 — Valve type decibel-meter

This type of decibel-meter usually consists of an amplifier followed by a detector stage that rectifies the amplified signal. The DC component of this rectified waveform is applied to a meter giving a deflection that may be made independent of the frequency of the incoming signal. The meter is calibrated directly in dbm. Damage to the meter from overloading is prevented by arranging that at least one stage of the amplifier acts as a limiter if too large an input signal is applied.

The input circuit may be arranged so that either LEVEL or TRANS measurements may be made. The range of the meter may be varied by adjusting the gain of the amplifier in fixed steps. Fig. 638 shows a decibel-meter of this type reading power level from — 20 to + 30 dbm in 5 ranges.

Measurement of losses and gains

There are two main methods of measuring the loss or gain of a network. The first is to apply a tone at the required frequency, and to determine the ratio of " power in " to " power out ", which,

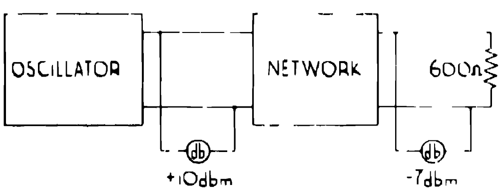


FIG 639 Measurement of the loss of a network by direct readings

using the decibel notation, will be the difference in the dbm readings at the input and the output. Thus in Fig. 639, if the input power is + 10 dbm and the output power — 7 dbm, the loss, or attenuation of the network, will be 17 db.

The second method is the method of substitution. This requires a calibrated variable attenuator. To measure a network having a loss, the tone is applied first through the network, and then

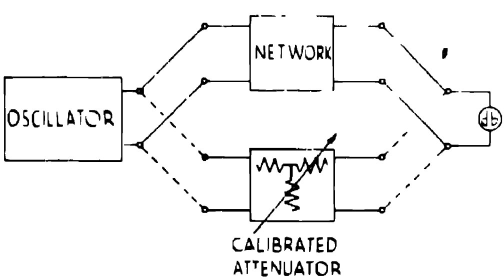


FIG. 640.—Measurement of the loss of a network by the method of substitution.

through the variable attenuator (Fig. 640). The attenuator is adjusted until the same power output level is obtained in the two cases, for constant input power. The known loss of the calibrated attenuator is then equal to the loss of the network.

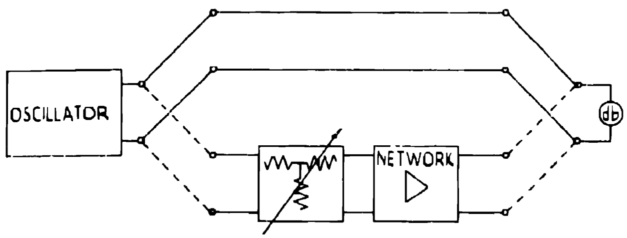


FIG. 641.—Measurement of the gain of a network.

To measure a gain, such as may occur if the network contains an amplifier, the method must be modified, since it is unlikely that a calibrated amplifier will be available. The calibrated variable attenuator may be utilised to determine the gain as follows :—

> The oscillator is first applied directly to the meter and the reading noted. The calibrated attenuator and the network are then inserted between the oscillator and the meter. The attenuator is adjusted until the same meter reading is obtained. When this is the case, the loss of the attenuator must equal the gain of the amplifier, which is therefore determined. The attenuator should be placed in the input to the amplifier, or the amplifier may be overloaded.

The advantage of this latter method over the former is that the results obtained are independent of the accuracy of the decibelmeter —in fact, an uncalibrated meter can be used, provided some method is employed to ensure that the same deflection is obtained in the two cases.

Transmission measuring sets

A transmission measuring set (TMS) is the name given to the apparatus used for measuring gains and losses. Provided that some source of test tone may be found, a simple decibel-meter may be used as a TMS by applying the first of the above methods.

In general, a TMS includes :—

(i) an oscillator ;

(ii) a means of calibrating the output of this oscillator ;

(iii) a means of measuring the incoming signal

Practical transmission measuring sets are usually designed on a 600 ohm basis, and this fact must be remembered when making measurements in circuits with an impedance different from this value.

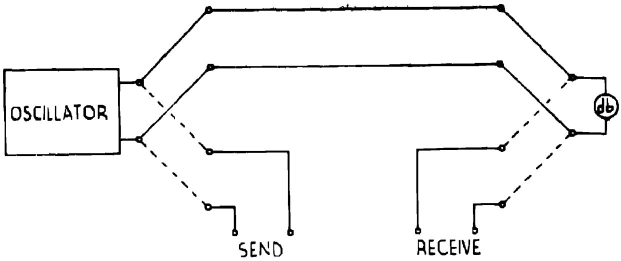


Fig 642 —Schematic arrangement of a TMS.

Fig 642 shows the schematic arrangement of such a TMS. To calibrate the oscillator output, the output is fed directly into the decibelmeter, and is adjusted to the required level. The output from the TMS is then applied to the circuit under test via the SEND terminal. The output from the circuit is re-applied to the TMS via the RECEIVE terminals, and the power level indicated on the meter. The difference between this latter reading and the oscillator output gives the loss or gain of the circuit under test. If the receive circuit has an input impedance of 600 ohms, only TRANS measurements may be made. Most TMSs, however, employ the principle previously discussed, *i.e.*, a high impedance voltmeter with a 600 ohms resistance in shunt if desired, thus enabling either TRANS or LEVEL measurements to be made.

The output of the oscillator may be varied over a wide range by incorporating an attenuator before the send terminals. A range of output from +20 dbm down to −50 dbm will enable most gains and losses to be measured.

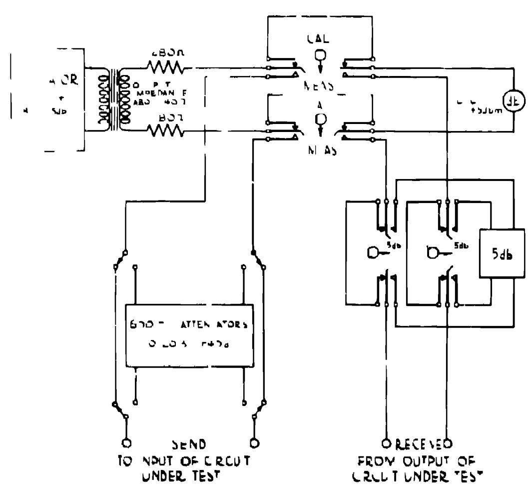


Fig. 643 —Transmission measuring set (" TRANS " measurements only)

Similar modification may be made to the receive circuit. A rectifier type meter may read directly from -10 to $+5$ dbm, but by incorporating a 15 db pad which may be inserted in the input before the meter, the upper limit of the readable level will be increased to $+20$ dbm.

Fig 643 shows a transmission measuring set having the oscillator output at the send terminals variable from $+5$ dbm down to -55 dbm, and capable of making TRANS measurements from $+20$ dbm down to -10 dbm.

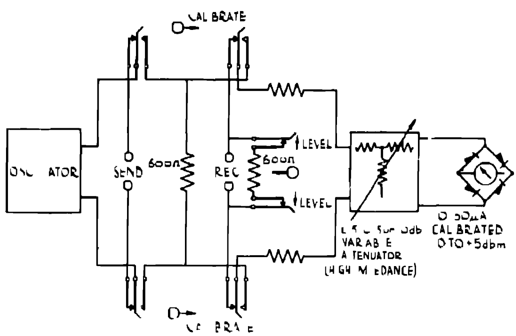


Fig. 644 —Transmission measuring set ("TRANS" and "LEVEL" measurements)

Fig 644 shows a transmission measuring set capable of making TRANS and LEVEL measurements on the receive side.

To measure the attenuation of a long transmission line, two TMSs are required, one to supply the standard level at the sending end, and the other to measure the power received at the distant end.

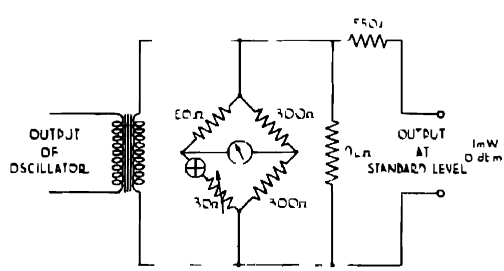


Fig 645 —Lamp bridge used for calibrating a TMS.

A "lamp bridge" provides a convenient method of calibrating a TMS, since it has a negligible frequency error. For example, it may be used as in Fig. 645 to ensure that the output from the

oscillator is exactly 1mW, even though the only meter available has a large frequency error The only disadvantage of the lamp bridge circuit is that it is difficult to make its impedance 600 ohms and a matching pad must be inserted between the bridge and the output terminals

The resistance of the lamp increases with the current through it The 30 ohm variable resistance is adjusted so that the bridge is balanced when the voltage applied across the bridge from the transformer has an RMS value of 1 49 volts This voltage applied to the subsequent network and 600 ohms load ensures an output of 1mW Therefore once the correct bridge setting has been obtained, the only requirement to ensure the output of 1mW is that the bridge shall be balanced i e nul deflection of the meter Since any frequency error of the meter will not affect the balance point of the bridge, exactly 1mW output will be obtained irrespective of the frequency

CHAPTER 15

FILTERS

A network that is designed to attenuate certain frequencies but pass others without loss is called a "filter." A filter therefore possesses at least one "pass band" (a band of frequencies in which the attenuation is zero) and at least one "attenuation band" (a band of frequencies in which the attenuation is finite). The frequencies that separate the various pass and attenuation bands are called "cut-off" frequencies and are usually denoted by f_1, f_2 *etc.*, or by f_0 if there is only a single cut-off frequency.

An important characteristic of all filters is that they are constructed from purely *reactive* elements, for otherwise the attenuation could never become zero. It is interesting to compare a filter, in which the attenuation changes *suddenly* from zero to some other value, with an attenuator pad (pure resistances only) of *constant* attenuation (independent of frequency), and with an equaliser (resistances and reactances) whose attenuation undergoes a *gradual* variation with frequency.

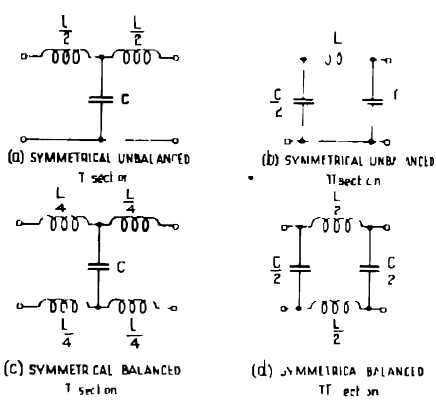


FIG. 646.—Symmetrical low-pass T and π sections in unbalanced and balanced forms.

The filters considered in this chapter will in general be symmetrical unbalanced T or π sections, as shown in Fig. 646*a* and *b* respectively, for the case of a low-pass filter.

Symmetrical balanced sections, as shown in Fig. 646*c* and *d*, will not be considered separately, since they may be deduced from the unbalanced sections, just as was done in the case of balanced attenuators.

Elementary filter sections

The simplest type of filter has only one pass band, one attenuation band, and a single cut-off frequency. If it passes all frequencies up to the cut-off frequency and attenuates all frequencies above, it is called a " low-pass " filter. If, on the other hand, it attenuates all frequencies below the cut-off frequency and passes all frequencies above, it is called a " high-pass " filter.

An " ideal " filter would have zero attenuation in the pass band and infinite attenuation in the attenuation band : an ideal low-pass filter, for example, might have an attenuation-frequency curve as shown in Fig. 647.

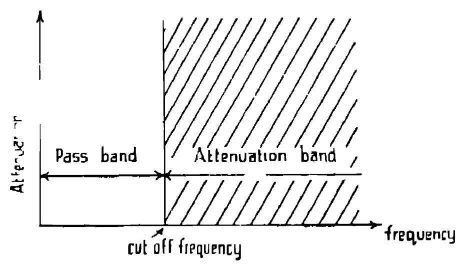


FIG. 647.—Attenuation-frequency curve of an ideal low-pass filter.

This cannot be achieved in practice. In the first place, in a practical filter it is found that the attenuation outside the pass band is finite ; it can, however, be made as large as required by using a sufficient number of sections in series. Secondly, if any resistance is present (and it is impossible to construct an inductance that does not possess a certain amount of resistance), the attenuation in the pass band will not be zero ; usually, however, it is only one or two decibels. Finally, mismatch losses must be considered ; for although the characteristic impedance of the section may vary with frequency, it will probably be terminated in a fixed resistance, or in an impedance that does not vary with frequency in the same way as the characteristic impedance of the section.

The complete study of the behaviour of any filter requires the calculation of its propagation constant (γ), attenuation (α), phase-shift (β) and characteristic impedance (Z_0) at any frequency, this involves a somewhat advanced mathematical treatment. It is possible, however, to find the pass and attenuation bands from an elementary consideration of the variation of Z_0 with frequency, and this method will be considered first.

Theorem connecting α and Z_0

It will be assumed that the filter is correctly terminated in its characteristic impedance ; the following theorem then applies :—

Over the range of frequencies for which the characteristic impedance Z_0 of a filter is purely resistive (real), the attenuation α is zero. Over the range of frequencies for which Z_0 is purely

reactive, the attenuation is greater than zero. The case where Z_0 is partly resistive and partly reactive cannot arise in purely reactive networks.

The validity of this theorem can be demonstrated from elementary considerations. If Z_0 is real, the filter and its termination will absorb *power* from any generator connected to it; as the filter is composed entirely of reactances it cannot itself absorb power, since in a reactance the current and voltage are always 90° out of phase. Hence *all* the power delivered by the generator must be passed through to the load and therefore there is no attenuation, *i.e.* $\alpha = 0$.

If, on the other hand, Z_0 is purely reactive, the filter and its termination cannot absorb any power, and no power is therefore passed to the load.

The last part of the above proof is rather unsatisfactory, and a more rigorous proof is therefore given below, using the normal formulae for a T or π section.

As any section may be represented by a simple T section, it is sufficient to consider just the T section shown in Fig. 648.

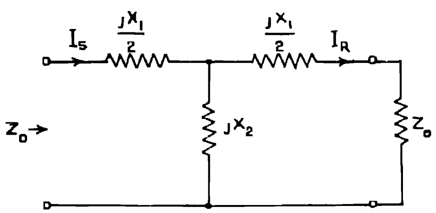


FIG. 648.—T section composed of pure reactive elements.

As the elements are all reactive, they may be written in the form jX, where X is real, but may be positive or negative.

The formulae for the characteristic impedance and propagation constant of a T section are:—

$$Z_0 = \sqrt{\frac{Z_1^2}{4} + Z_1Z_2} \quad \text{(Equation 10 of Chap. 13)} \qquad (1)$$

and
$$e^{\gamma} = \frac{I_S}{I_R} = 1 + \frac{Z_1}{2Z_2} + \frac{Z_0}{Z_2} \quad \text{(Equation 16 of Chap. 13)} \qquad (2)$$

and α can be calculated from the relationship:—

$$\alpha = 20 \log_{10} \left| \frac{I_S}{I_R} \right| \text{ db} \qquad (3)$$

In this case $Z_1 = jX_1$ and $Z_2 = jX_2$, and equations 1 and 2 become:—

$$Z_0 = j\sqrt{\frac{X_1^2}{4} + X_1X_2} \qquad (4)$$

and
$$\frac{I_S}{I_R} = 1 + \frac{X_1}{2X_2} - j\frac{Z_0}{X_2} \qquad (5)$$

There are two cases to consider, depending on the value and sign of X_1 and X_2. The two cases are :—

$$(a) \qquad \frac{X_1^2}{4} + X_1X_2 \quad \text{is negative,} \; = -A, \text{ say,}$$

$$\text{and} \quad (b) \qquad \frac{X_1^2}{4} + X_1X_2 \quad \text{is positive,} \; = +B, \text{ say,}$$

where A and B are real and positive.

Consider first case (a). In this case, from equation 4, Z_0 is real (that is, purely resistive), and its value is :—

$$Z_0 = j\sqrt{\frac{X_1^2}{4} + X_1X_2} = j\sqrt{-A} = +\sqrt{A}$$

From equation 5 :—

$$\frac{I_S}{I_R} = \left[1 + \frac{X_1}{2X_2}\right] - j\left[\frac{Z_0}{X_2}\right] = \left[1 + \frac{X_1}{2X_2}\right] - j\left[\frac{\sqrt{A}}{X_2}\right]$$

$$\therefore \quad \left|\frac{I_S}{I_R}\right| = \sqrt{\left(1 + \frac{X_1}{2X_2}\right)^2 + \frac{A}{X_2^2}}$$

$$= \sqrt{1 + \frac{X_1}{X_2} + \frac{X_1^2}{4X_2^2} - \frac{X_1^2}{4X_2^2} - \frac{X_1}{X_2}}$$

$$= 1$$

$$\text{But} \qquad \alpha = 20 \log_{10} \left|\frac{I_S}{I_R}\right|$$

Hence $\quad \alpha = 0$ if Z_0 is real.

In case (b), Z_0 is imaginary (that is, purely reactive), and its value, from equation 4, is :—

$$Z_0 = j\sqrt{\frac{X_1^2}{4} + X_1X_2} = j\sqrt{+B}$$

In this case, from equation 5 :—

$$\frac{I_S}{I_R} = \left[1 + \frac{X_1}{2X_2}\right] - j\left[\frac{Z_0}{X_2}\right] = \left[1 + \frac{X_1}{2X_2}\right] - j\left[\frac{j\sqrt{B}}{X_2}\right]$$

$$\therefore \quad \left|\frac{I_S}{I_R}\right| = 1 + \frac{X_1}{2X_2} + \frac{\sqrt{\frac{X_1^2}{4} + X_1X_2}}{X_2}$$

which is real and cannot be unity.

Therefore $\alpha \neq 0$ if Z_0 is imaginary.

Determination of cut-off frequency

The theorem just given is of fundamental importance, since it can be applied to determine the cut-off frequency f_c of any filter, from a consideration of Z_0. For Z_0 is real in a pass band and imaginary in an attenuation band; hence *f_c is the frequency at which Z_0 changes from being real to being imaginary.*

This point can easily be found, for a T section, by using the formula :—

$$Z_0 = \sqrt{\frac{Z_1^2}{4} + Z_1 Z_2}$$

If all the elements are pure reactances, let $Z_1 = jX_1$ and $Z_2 = jX_2$, so that :—

$$Z_0 = \sqrt{\frac{-X_1^2}{4} - X_1X_2} = \sqrt{-X_1\left(\frac{X_1}{4} + X_2\right)} \quad (6)$$

Hence if X_1 and $\frac{X_1}{4} + X_2$ have the *same* sign, Z_0 will be the square root of a negative quantity, *i.e., purely imaginary*, and the filter will attenuate. If, however, X_1 and $\frac{X_1}{4} + X_2$ have *opposite* signs, Z_0 will be *real* and the attenuation zero.

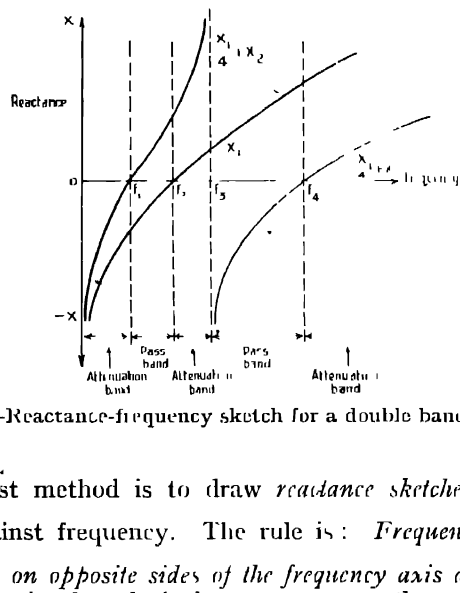


FIG. 649 —Reactance-frequency sketch for a double band-pass filter.

The easiest method is to draw *reactance sketches* for X_1 and $\frac{X_1}{4} + X_2$ against frequency. The rule is : *Frequencies for which the curves are on opposite sides of the frequency axis are in the pass band ; frequencies for which the curves are on the same side of the frequency axis are in the attenuation band ; the change-over points give the cut-off frequencies.* For example, the filter whose reactance sketches are shown in Fig. 649 has two pass bands and three attenuation bands. An alternative method is to write down Z_0 in terms of frequency and calculate algebraically where it is resistive and where it is reactive.

Constant-*k* sections

Before applying these methods, the following definition is required :—

A " constant-*k* " section is a T or π section in which the

series and shunt impedances, Z_1 and Z_2, are connected by the relationship $Z_1Z_2 = R_0^2$, where R_0 is a real constant—that is, a resistance that is independent of frequency. R_0 is known as the "design impedance" of the section.

Consider the equation :—

$$Z_{0T} = \frac{Z_1 Z_2}{Z_{0\pi}} \quad . \tag{7}$$

which connects the characteristic impedances of T and π sections composed of the same series and shunt impedances (*see* Fig. 650). This formula is proved in Chapter 13, on page 578.

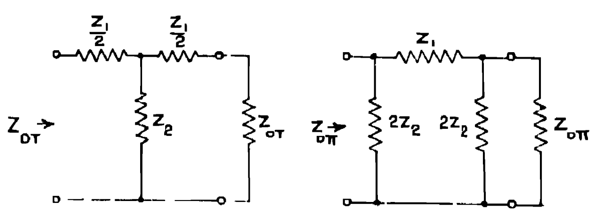


FIG. 650 —T and π sections composed of the same series and shunt impedances.

If the section is a constant-k section :—

$$Z_{0\pi} = \frac{R_0^2}{Z_{0T}} \tag{8}$$

and clearly $Z_{0\pi}$ and Z_{0T} will be real or imaginary together; and when Z_{0T} changes from real to imaginary, so also will $Z_{0\pi}$. Hence the two sections will have the same pass bands and the same cut-off frequencies.

The constant-k T or π sections of any type of filter are known as the *prototypes*; other more complex sections may be derived from the prototype, and these will be dealt with later. At the moment, the low-pass and high-pass prototype sections will be considered.

PROTOTYPE FILTER SECTIONS

Low-pass filters

The prototype T and π low-pass filter sections are shown in Fig. 651.

Here $\qquad Z_1 = j\omega L$

and $\qquad Z_2 = \frac{-j}{\omega C}$

Hence $Z_1Z_2 = \frac{L}{C}$, and the sections are therefore constant-k sections with :—

$$R_0 = \sqrt{\frac{L}{C}} \tag{9}$$

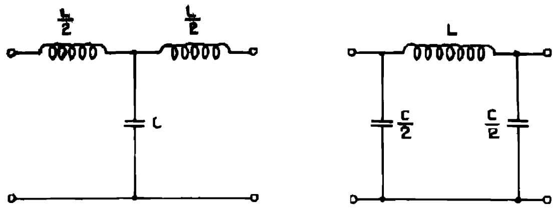


FIG 651 —Prototype T and π low pass filter sections

As both sections have the same cut-off frequency, it is sufficient to calculate this for the T section only The reactance sketches must first be drawn —

$$Z_1 = j\omega L \qquad \therefore\ X_1 = \omega L$$

$$Z_2 = \frac{-j}{\omega C} \qquad \therefore\ X_2 = \frac{-1}{\omega C}$$

$$\therefore \qquad \frac{X_1}{4} + X_2 = \frac{\omega L}{4} - \frac{1}{\omega C} \tag{10}$$

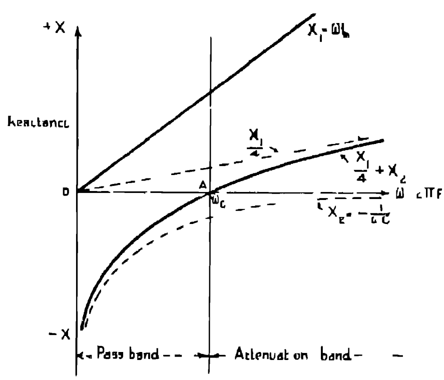


FIG 652 —Reactance frequency sketch for a prototype low-pass filter

The curves are as shown in Fig 652 It is worth noting here that all reactance-frequency curves slope upwards and to the right; that is, they have positive slope

It will be seen that the curves are on opposite sides of the frequency axis as far as the point A, and on the same side from A onwards Hence the pass band includes all frequencies up to the point A, and the attenuation band all frequencies above the point A. The point A itself marks the cut-off frequency given by $\omega = \omega_0$. The section is therefore a low-pass filter with a cut-off frequency

$$f_0 = \frac{\omega_0}{2\pi}.$$

ω_0 is the point where the curve $\frac{X_1}{4} + X_2$ crosses the frequency axis; that is, where $\frac{\omega L}{4} - \frac{1}{\omega C} = 0$ or $\omega^2 = \frac{4}{LC}$.

Hence $$\omega_0 = \frac{2}{\sqrt{LC}} \text{ or } f_0 = \frac{1}{\pi\sqrt{LC}} \quad (11)$$

which gives the cut-off frequency of a low-pass T or π section.

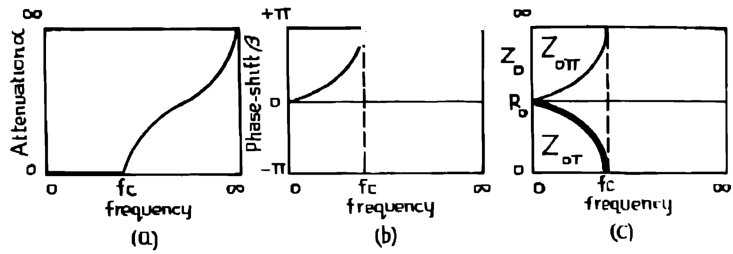


FIG. 653 —Prototype low-pass filter sections—Variation with frequency of

(a) Attenuation (b) Phase-shift (c) Characteristic impedance

Fig. 653 shows the way in which α, β and Z_0 vary with frequency for a prototype low-pass section; Z_0 is not shown above the cut-off frequency as it becomes reactive in the attenuation band. It should be noted that Z_0 varies considerably from the design impedance R_0 over the pass band.

The algebraic approach to the same problem is as follows: –

$$Z_{0T} = \sqrt{\frac{Z_1^2}{4} + Z_1 Z_2}$$

$$= \sqrt{\frac{-\omega^2 L^2}{4} + \frac{L}{C}}$$

$$= \sqrt{\frac{L}{C}} \times \sqrt{1 - \frac{\omega^2 LC}{4}}$$

i.e. $$Z_{0T} = R_0\sqrt{1 - \frac{\omega^2 LC}{4}} \quad (12)$$

Clearly Z_{0T} is real if $\frac{\omega^2 LC}{4} < 1$ and imaginary if $\frac{\omega^2 LC}{4} > 1$; hence the section passes frequencies below $\omega = \frac{2}{\sqrt{LC}}$ and attenuates frequencies above this value. It is therefore a low-pass filter with a cut-off frequency given by $\omega_0 = \frac{2}{\sqrt{LC}}$ or $f_0 = \frac{1}{\pi\sqrt{LC}}$; this agrees with equation 11, found by the reactance sketch method.

Note that equation 12 may be written as :—

$$Z_{0T} = R_0\sqrt{1 - \frac{\omega^2}{\omega_c^2}} \qquad (13)$$

Since $Z_{0T} = \dfrac{R_0^2}{Z_{0\pi}}$ (from equation 8), it follows that :

$$Z_{0\pi} = \frac{R_0}{\sqrt{1 - \dfrac{\omega^2}{\omega_c^2}}} \qquad (14)$$

From these two expressions the curves of Fig 653c are derived

Example —

Consider a simple T section low-pass filter having a design impedance R_0 Find $Z_{0\pi}$ at $f = 0{\cdot}9 f_c$.

$$\frac{\omega}{\omega_c} = \frac{f}{f_c} = 0{\cdot}9$$

$$\therefore \qquad Z_{0\pi} = \frac{R_0}{\sqrt{1 - (0{\cdot}9)^2}} = 2{\cdot}3\ R_0$$

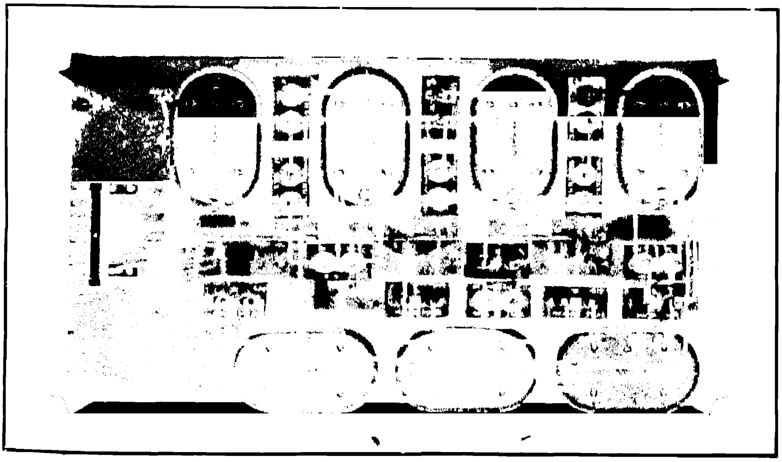


PLATE 33 —Low pass filter on a multi-channel carrier telephone system.

More advanced mathematical treatment of the low-pass filter

In order to determine the phase-shift inside the pass band a the attenuation outside the pass band, it is convenient to introdu hyperbolic functions It has been seen (equation 20, p 573) th for a T section :—

$$\cosh \gamma = 1 + \frac{Z_1}{2Z_2}$$

where γ is the propagation constant of the section Applying t to the case of the low-pass filter gives —

$$\cosh \gamma = 1 - \frac{\omega^2 LC}{2} \tag{15}$$

Since $\gamma = \alpha + j\beta$, it follows that —

$$\cosh \alpha \cos \beta + j \sinh \alpha \sin \beta = 1 - \frac{\omega^2 LC}{2}$$

Equating real and imaginary parts —

$$\cosh \alpha \cos \beta = 1 - \frac{\omega^2 LC}{2} \tag{16}$$

and

$$\sinh \alpha \sin \beta = 0 \tag{17}$$

Equation 17 is satisfied by —

$$\text{either} \quad \alpha = 0 \quad \text{or} \quad \beta = n\pi$$

Pass band

If $\alpha = 0$, equation 16 gives :—

$$\cos \beta = 1 - \frac{\omega^2 LC}{2} \tag{18}$$

This corresponds to the pass band of the filter The attenuation i zero and the phase shift is given by equation 18 Since cos β must lie between + 1 and — 1, equation 18 can be satisfied for all frequencies from zero up to the cut off frequency f_0, where —

$$1 - \frac{\omega_0^2 LC}{2} = -1$$

i e

$$\frac{\omega_0^2 LC}{2} = 2$$

i e.

$$\omega_0 = \frac{2}{\sqrt{LC}} \tag{19}$$

i e.

$$f_0 = \frac{1}{\pi\sqrt{LC}} \tag{20}$$

which verifies the result already obtained

From equation 19, equation 18 becomes —

$$\cos \beta = 1 - 2\frac{\omega^2}{\omega_0^2} \tag{21}$$

Hence

$$\beta = \cos^{-1}\left(1 - 2\frac{\omega^2}{\omega_0^2}\right) \text{ radians} \tag{22}$$

Alternatively to equation 21 :—

$$1 - 2\sin^2\frac{\beta}{2} = 1 - 2\frac{\omega^2}{\omega_o^2}$$

Whence $$\beta = 2\sin^{-1}\frac{\omega}{\omega_o} \text{ radians.} \qquad (23)$$

Attenuation Band

If $\beta = n\pi$, then $\cos\beta = \pm 1$, and equation 16 gives :—

$$\pm\cosh\alpha = 1 - \frac{\omega^2 LC}{2}$$

From equation 19 this becomes :—

$$\pm\cosh\alpha = 1 - 2\frac{\omega^2}{\omega_o^2} \qquad (24)$$

The attenuation band of the low-pass filter is given by :—

$$\beta = \pi$$

and $$\cosh\alpha = 2\frac{\omega^2}{\omega_o^2} - 1$$

In the attenuation band :—

$$\alpha = \cosh^{-1}\left(\frac{2\omega^2}{\omega_o^2} - 1\right) \text{ nepers} \qquad (25)$$

or $$\alpha = 2\cosh^{-1}\frac{\omega}{\omega_o} \text{ nepers} \qquad (26)$$

These results give the curves of Fig 653*a* and *b*

Example :—

At what frequency will a T section low-pass filter, having a cut-off frequency f_o, have an attenuation of 10 db ?

$$10 \text{ db} \equiv 1\cdot15 \text{ neper}$$

Let $\omega = 2\pi f$, where f is the required frequency.

Then $$2\cosh^{-1}\frac{\omega}{\omega_c} = 1\cdot15$$

$$\therefore \frac{f}{f_o} = \frac{\omega}{\omega_o} = \cosh 0\cdot575 \quad (\textit{see } \text{p. } 796)$$

$$\therefore f = 1\cdot17 f_c$$

Design of a prototype low-pass filter section

Let the design impedance R_0 and the cut-off frequency f_o be given There are two equations giving these in terms of L and C, namely :—

$$R_0 = \sqrt{\frac{L}{C}} \qquad \text{(from equation 9)}$$

and $$f_o = \frac{1}{\pi\sqrt{LC}} \qquad \text{(from equation 11)}$$

$$\text{Whence} \qquad L = \frac{R_0}{\pi f_0} \qquad (27)$$

$$\text{and} \qquad C = \frac{1}{\pi R_0 f_0} \qquad (28)$$

From these equations the components of the required section can be determined

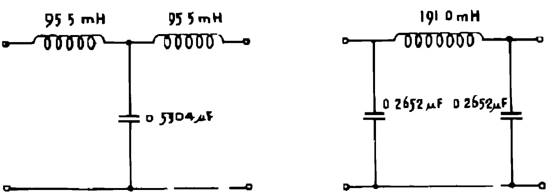
Example —

Suppose a prototype low-pass filter section is required having $R_0 = 600\,\Omega$ and $f_0 = 1000$ c/s Substituting the known values of R_0 and f_0 in equations 27 and 28,

$$L = \frac{R_0}{\pi f_0} = \frac{600}{1000\pi} = 191\ 0 \text{ mH}$$

and

$$C = \frac{10^6}{\pi\ 1000\ 600}\ \mu\text{F} = 0\ 5304\ \mu\text{F}$$



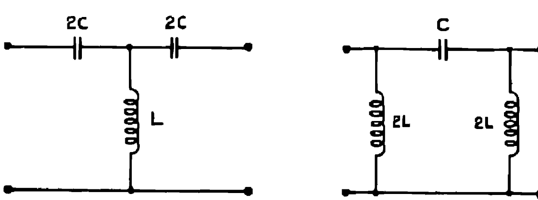
(a) (b)

FIG 654

Hence the prototype T section is as shown in Fig 654*a* and the π section as shown in Fig 654*b* Each has a cut-off frequency of 1000 c/s and a design impedance 600 Ω the difference between the two is in the way in which the characteristic impedance varies with frequency

High-pass filters

Fig 655 shows the prototype high-pass T and π sections



(a) T SECTION (b) π SECTION

FIG 655 —Prototype high-pass filter sections

$Z_1 = \frac{-j}{\omega C}$, and $Z_2 = j\omega L$. Thus $Z_1 Z_2 = \frac{L}{C}$, and the sections are constant-k sections with :—

$$R_0 = \sqrt{\frac{L}{C}} \tag{29}$$

The cut-off frequency may be determined by the reactance sketch method. In this case :—

$$X_1 = -\frac{1}{\omega C} \text{ and } X_2 = \omega L$$

$$\therefore \qquad \frac{X_1}{4} + X_2 = \omega L - \frac{1}{4\omega C} \tag{30}$$

The curves are as shown in Fig. 656.

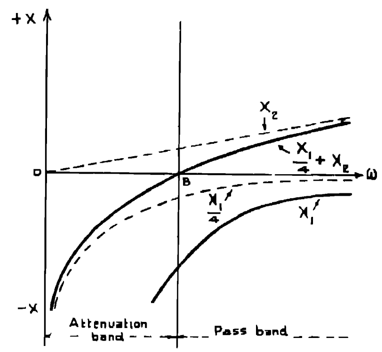


FIG 656 —Reactance-frequency curves for a prototype high-pass filter.

Here the curves are on the same side of the horizontal axis up to the point B, giving an attenuation band. For frequencies above B, the curves are on opposite sides of the axis, giving a pass band. The point B therefore gives the cut-off frequency. This frequency is given by :—

$$\frac{X_1}{4} + X_2 = 0$$

i.e.

$$\omega L - \frac{1}{4\omega C} = 0$$

Hence

$$\omega_0 = \frac{1}{2\sqrt{LC}} \text{ or } f_0 = \frac{1}{4\pi\sqrt{LC}} \tag{31}$$

which gives the cut-off frequency of a high-pass T or π section.

Fig. 657 shows the way in which α, β and Z_0 vary with frequency for a prototype high-pass section. Z_0 is not shown below the cut-off frequency, as it becomes reactive in the attenuation band.

The algebraic approach is as follows :—

$$Z_{0T} = \sqrt{\frac{Z_1^2}{4} + Z_1Z_2} = \sqrt{\frac{-1}{4\,\omega^2C^2} + \frac{L}{C}} = \sqrt{\frac{L}{C}} \times \sqrt{1 - \frac{1}{4\,\omega^2 LC}}$$

$$Z_{0T} = R_0\sqrt{1 - \frac{1}{4\,\omega^2LC}} \qquad (32)$$

It follows that if $4\,\omega^2LC > 1$, Z_{0T} is real and the filter passes. If $4\,\omega^2LC < 1$, Z_{0T} is imaginary and the filter attenuates.

Hence the cut off frequency is given by :-

$$4\,\omega^2LC = 1$$

i e $$\omega_0 = \frac{1}{2\sqrt{LC}}$$

i e $$f_0 = \frac{1}{4\pi\sqrt{LC}}$$ (which agrees with equation 31)

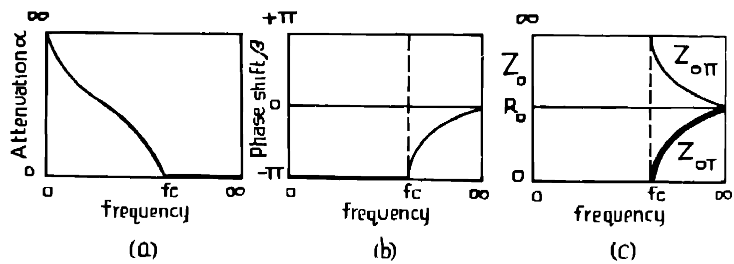


Fig 657 —Prototype high pass filter—Variation with frequency of

(a) Attenuation (b) Phase shift (c) Characteristic impedance

Note that Z_{0T} may be written as

$$Z_{0T} = R_0\sqrt{1 - \frac{\omega_0^2}{\omega^2}} \qquad (33)$$

and $$Z_{0\pi} = \frac{R_0^2}{Z_{0T}} = \frac{R_0}{\sqrt{1 - \frac{\omega_0^2}{\omega^2}}} \qquad (34)$$

from which results the curves of Fig 657c are obtained

It can also be shown that, outside the pass band :—

$$\alpha = \cosh^{-1}\left(\frac{2\,\omega_0^2}{\omega^2} - 1\right) = 2\cosh^{-1}\frac{\omega_0}{\omega}\ \text{nepers} \qquad (35)$$

and within the pass band :-

$$\beta = \cos^{-1}\left(1 - \frac{2\,\omega_0^2}{\omega^2}\right) = -2\sin^{-1}\frac{\omega_0}{\omega}\ \text{radians} \qquad (36)$$

These results give the curves of Fig. 657*a* and *b*.

It will be noted that the formulae 13, 14, 26 and 23 for a low-pass filter correspond exactly with the formulae 33, 34, 35 and 36 for a high-pass filter if $\frac{\omega_0}{\omega}$ is written for $\frac{\omega}{\omega_0}$.

Design of a prototype high-pass filter section

Let the design impedance R_0 and the cut-off frequency f_0 be given. There are two equations giving these in terms of L and C, namely :—

$$R_0 = \sqrt{\frac{L}{C}} \qquad \text{(from equation 29)}$$

and

$$f_0 = \frac{1}{4\pi\sqrt{LC}} \qquad \text{(from equation 31)}$$

Whence

$$L = \frac{R_0}{4\pi f_0} \tag{37}$$

and

$$C = \frac{1}{4\pi R_0 f_0} \tag{38}$$

From these equations, the component values of the required section can be determined.

Example.—

Calculate the components of a prototype high-pass filter section having a design impedance $R_0 = 600\Omega$ and a cut-off frequency $f_0 = 10$ kc/s.

Substituting the known values of R_0 and f_0 in equations 37 and 38 :—

$$L = \frac{600}{4\pi \times 10^4} = 4 \cdot 774 \text{ mH}$$

and

$$C = \frac{10^6}{4\pi\ 600 \times 10^4} = 0 \cdot 01326\ \mu\text{F}$$

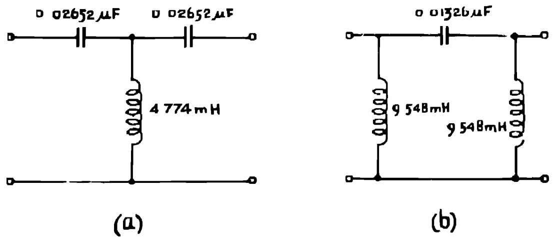


FIG. 658.

Hence the prototype T section is as shown in Fig. 658*a* and the π section as shown in Fig. 658*b*.

M-DERIVED FILTER SECTIONS

Behaviour of prototype sections

There are two obvious disadvantages of the prototype sections just discussed. In the first place, considering a low-pass section the attenuation does not rise very rapidly after the cut-off frequency, being only 10 db at a frequency $f = 1 \cdot 2 f_c$; secondly, Z_0 is by no means constant over the pass band (*e.g.* $Z_{0\pi} = 2 \cdot 3\ R_0$ at $f = 0 \cdot 9 f_c$).

Consider first the attenuation band; the most obvious way to increase the attenuation beyond cut-off is by connecting two or more sections together. This can be done provided the impedances match correctly—*e.g.* a T and a π section cannot be connected together since $Z_{0T} \neq Z_{0\pi}$, but two or more T sections may be connected in series, as also may two or more π sections. Thus with two sections in series, although the attenuation over the pass band is, in theory, still zero, the attenuation over the attenuation band is doubled; so that, for example, with two low-pass filter sections, the attenuation at $f = 1 \cdot 2 f_c$ is 20 db, showing a much sharper cut-off than that obtained with a single section.

Unfortunately, due to resistance in the components, the attenuation in the pass band of a practical filter is not zero, but becomes quite appreciable towards cut-off. The result is that the

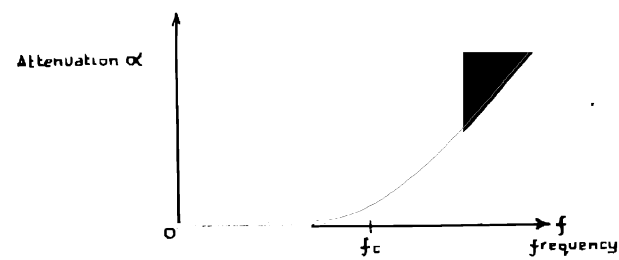


FIG. 659.—Attenuation-frequency characteristic of a typical prototype low-pass filter.

curve becomes rounded off at the cut-off frequency (*see* Fig. 659, which shows the attenuation-frequency characteristic of a typical low-pass filter).

What is really required is a section having the same cut-off frequency as the prototype section, but a different attenuation-frequency characteristic in the attenuation band; that is, a characteristic that rises more rapidly than that of the prototype. It will also be necessary for the new section to have the same Z_0 as the prototype *at all frequencies*, since otherwise the two sections cannot be connected together without mismatch. It might be thought impossible to satisfy both these requirements, but it will now be shown that it can be accomplished quite simply. Note that if the two sections have the same Z_0, they must also have the same pass bands.

Derivation of m-derived sections

Consider first any T section, and construct a new section from it, having a series arm of the same type but of different value ; *i.e.* for convenience make the new Z_1 equal to mZ_1, where m is some constant.

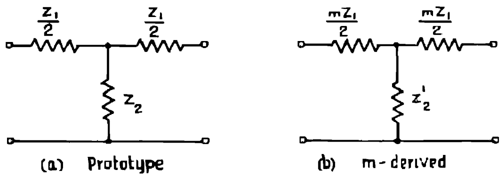


FIG. 660.—Derivation of m-derived T section

The new shunt arm will be not Z_2 but Z'_2, say, and it is necessary to find that value of Z'_2 (if any) which will make the two sections have the same value for Z_0.

For the prototype section :

$$Z_{0T} = \sqrt{\frac{Z_1^2}{4} + Z_1Z_2} \tag{39}$$

For the new section :

$$Z_{0T} = \sqrt{\frac{m^2Z_1^2}{4} + mZ_1Z'_2} \tag{40}$$

These two impedances will be the same if :

$$\frac{Z_1^2}{4} + Z_1Z_2 = \frac{m^2Z_1^2}{4} + mZ_1Z_2'$$

i.e. if

$$Z'_2 = \frac{Z_2}{m} + \frac{1-m^2}{4m} Z_1 \tag{41}$$

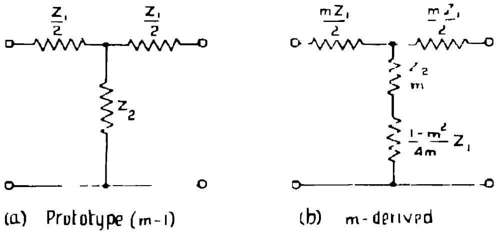


FIG. 661.—Comparison of prototype and m-derived T sections.

This means that Z'_2 must be an impedance $\frac{Z_2}{m}$ in series with an impedance $Z_1 \frac{(1-m^2)}{4m}$, and both these impedances can be constructed if $0 < m < 1$. The complete "m-derived T section" is shown in Fig. 661*b*.

In a similar manner, a new section may be derived from π section, having the same Z_0 at all frequencies. The complete m-derived π section is shown in Fig. 662*b*.

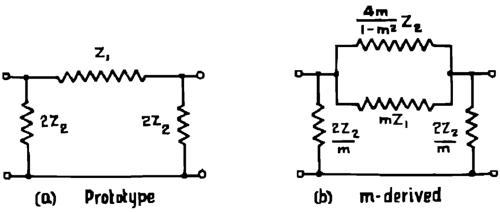


FIG. 662 —Comparison of prototype and m-derived π sections.

Note that if $m = 1$, both T and π m-derived sections reduce to the corresponding prototype sections. Note also that the characteristic impedance of the m-derived section is the same as that of the prototype, *i.e.* Z_{0T} or $Z_{0\pi}$.

The m-derived sections considered so far are perfectly general; to study further their behaviour, particular cases must be taken individually.

Low-pass m-derived sections

Fig. 663 shows both the m-derived T and π low-pass filter sections. In deducing these from the general case, note that to divide the impedance of a condenser by m, its capacity must be multiplied by m.

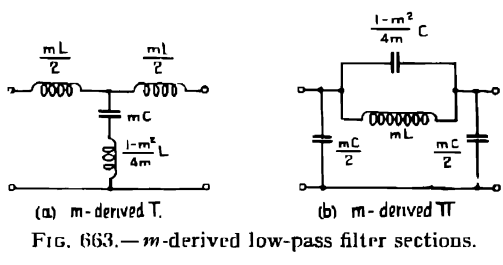


FIG. 663.—m-derived low-pass filter sections.

One important deduction can at once be made, namely, considering first the m-derived T section, that at some frequency the shunt arm will have a series resonance, giving a short-circuit across the transmission path and hence infinite attenuation. In the prototype sections, on the other hand, the attenuation becomes infinite only at infinite frequency. The frequency of infinite attenuation is denoted by f_r, and is given for an m-derived T section by the series resonance of the shunt arm.

This frequency is given by :—

$$\omega^2 = \frac{1}{\frac{1-m^2}{4m}L \times mC} = \frac{4}{LC\,(1-m^2)} = \frac{\omega_c^2}{1-m^2}$$

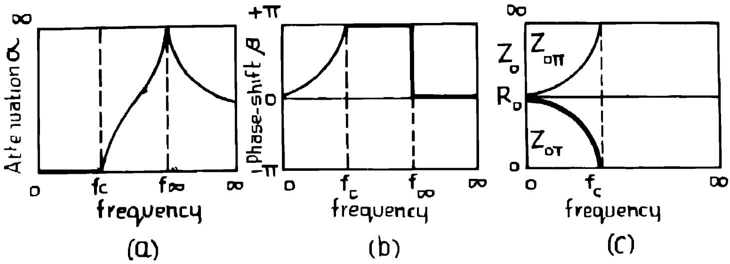


FIG. 664.—*m*-derived T and π low-pass filter sections—Variation with frequency of
(*a*) Attenuation (*b*) Phase-shift (*c*) Characteristic impedance

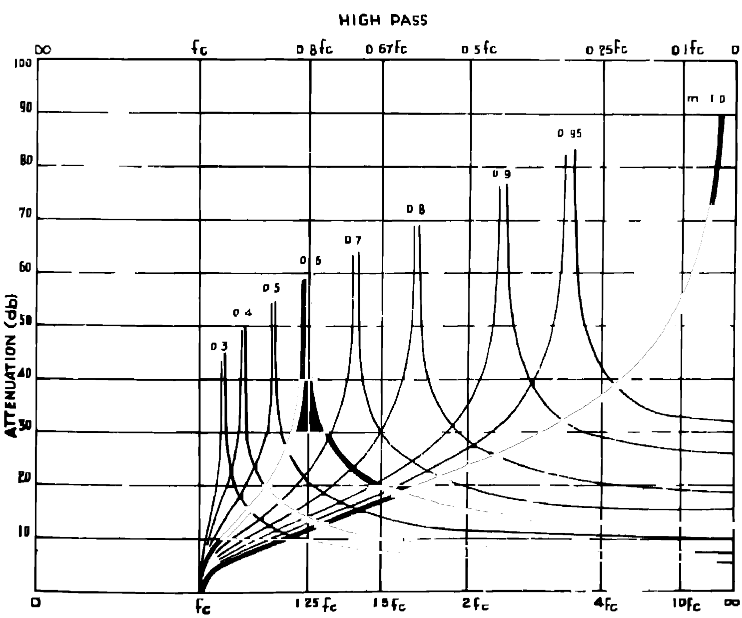


FIG. 665.—Attenuation frequency curves for *m*-derived T and π low-pass filter sections showing the effect of the value of *m*

Thus $\quad m_{\infty} = \dfrac{\omega_0}{\sqrt{1-m^2}} \quad$ or $\quad f_{\infty} = \dfrac{f_c}{\sqrt{1-m^2}} \qquad (42)$

Similarly an m-derived π section gives a frequency of infinite attenuation at the anti-resonance of the series arm; clearly this frequency has the same value as that for the T section. Fig. 664 shows the way in which α, β and Z_0 vary with frequency for an m-derived low-pass section.

Fig. 665 shows the attenuation-frequency characteristics for m-derived sections for various values of m. It will be seen that, particularly for small values of m, the attenuation rises much more rapidly than in the prototype, but falls off again after the frequency of infinite attenuation has been passed.

If f_∞ and f_0 are known, the required value of m may be calculated from equation 42 :—

$$\frac{f_\infty}{f_0} = \frac{1}{\sqrt{1-m^2}}$$

$$\therefore \qquad 1 - m^2 = \frac{f_0^2}{f_\infty^2}$$

$$\therefore \qquad m = \sqrt{1 - \frac{f_0^2}{f_\infty^2}} \qquad (43)$$

This result is used when designing a low-pass filter section to have an infinite attenuation at a given frequency f_∞.

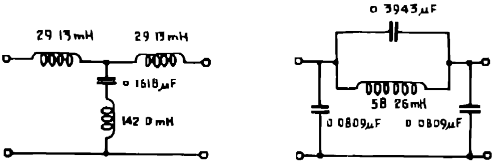
Example.—

Find m-derived T and π low-pass filter sections having a cut-off frequency $f_0 = 1000$ c/s, a design impedance $R_0 = 600\,\Omega$, and a frequency of infinite attenuation $f_\infty = 1050$ c/s.

The value of m is given by :—

$$m = \sqrt{1 - \frac{f_0^2}{f_\infty^2}} = \sqrt{1 - \left(\frac{1000}{1050}\right)^2} = 0{\cdot}305$$

It has already been shown that $L = 191{\cdot}0$mH and $C = 0{\cdot}5304\mu$F for the prototype of this filter (*see* example on page 646). The required m-derived sections are therefore as shown in Fig. 666.



(a) m-derived T section (b) m-derived π section

FIG. 666.

High-pass m-derived sections

Fig. 667 shows the T and π m-derived high-pass filter sections.

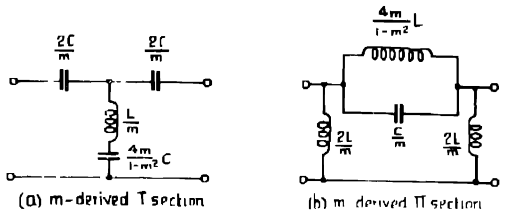


FIG. 667.—m-derived high-pass filter sections.

These have a frequency of infinite attenuation given by the resonance of the shunt arm in the T section, and by the anti-resonance of the series arm in the π section. This frequency is the same for the two sections, and is given by :—

$$\omega_{\infty}^2 = \frac{1 - m^2}{4LC} = (1 - m^2)\,\omega_c^2$$

$$\therefore \qquad \omega_{\infty} = \omega_c\sqrt{1 - m^2} \quad \text{or} \quad f_{\infty} = f_c\sqrt{1 - m^2} \qquad (44)$$

Fig. 668 shows the way in which α, β and Z_0 vary with frequency for an m-derived high-pass section.

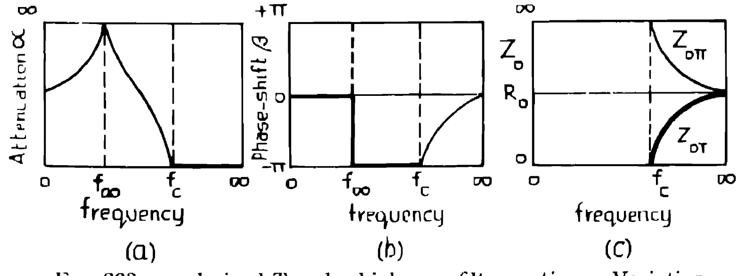


FIG. 668.—m-derived T and π high-pass filter sections—Variation with frequency of

(a) Attenuation. (b) Phase shift (c) Characteristic impedance.

The effect of different values of m on the attenuation-frequency characteristics corresponds exactly to that for the low pass filter if $\frac{f_c}{f}$ is written in place of $\frac{f}{f_c}$. Fig. 665 is therefore applicable to a high-pass filter provided that the top frequency scale is used. It follows that, as with the m-derived low-pass filter, a small value of m gives a very sharp cut-off.

If f_∞ and f_c are known, the value of m can be calculated from equation 44 :—

$$f_\infty = f_c\sqrt{1 - m^2}$$

$$\therefore \qquad m = \sqrt{1 - \left(\frac{f_\infty}{f_c}\right)^2} \qquad (45$$

This result is used when designing a high-pass filter to have an infinite attenuation at a given frequency f_∞.

Example.—

Design m-derived T and π high-pass filter sections having a cut-off frequency $f_c = 10$ kc/s, design impedance $R_0 = 600\Omega$ and $m = 0 \cdot 35$, and find the frequency of infinite attenuation.

This gives :— $f_\infty = f_c\sqrt{1 - m^2} = 10^4\sqrt{1 - 0\cdot35^2} = 9367$ c/s

It has already been shown that $L = 4 \cdot 774$mH and $C = 0 \cdot 01326\mu$F for the prototype of this filter (*see* example on page 649).

From Fig. 667 the m-derived sections having $m = 0 \cdot 35$ may be determined.

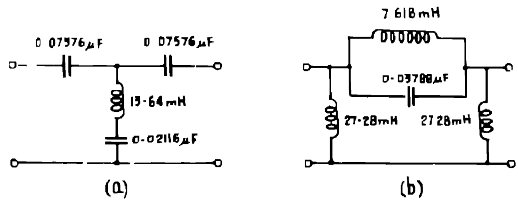


Fig. 669.

The required m-derived sections are as shown in Fig. 669.

IMPEDANCE MATCHING OF FILTERS

Impedance-matching half-sections

When a filter is composed of a number of sections, it is essential that the impedances at each junction shall be correctly matched.

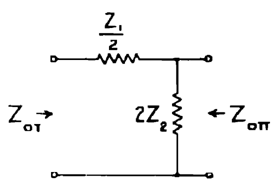


Fig. 670.—Prototype half section, showing image impedances.

Thus a T section having an impedance Z_{0T} should not be joined to a π section having an impedance $Z_{0\pi}$. If it is desirable to construct a filter containing both T and π sections, matching half-sections should be used.

The half-section shown in Fig. 670 has image impedances Z_{0T} and $Z_{0\pi}$, and may therefore be employed to match a T section to a π section as in Fig. 671.

If a filter is correctly matched throughout on such an image impedance basis, the overall attenuation of the filter will be simply

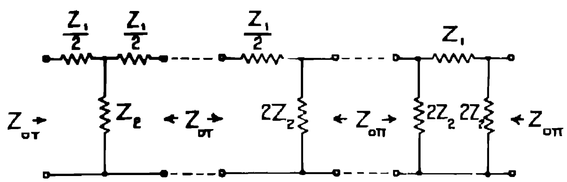


FIG. 671.—Use of prototype half-section for matching T and π sections.

the sum of the attenuations of the individual sections or half sections, there being no internal reflection or mismatch losses.

It will be remembered that the half-section is the basic element of the symmetrical ladder networks, and that both T and π sections when divided down the centre give rise to two of these half-sections (*see* Fig. 672).

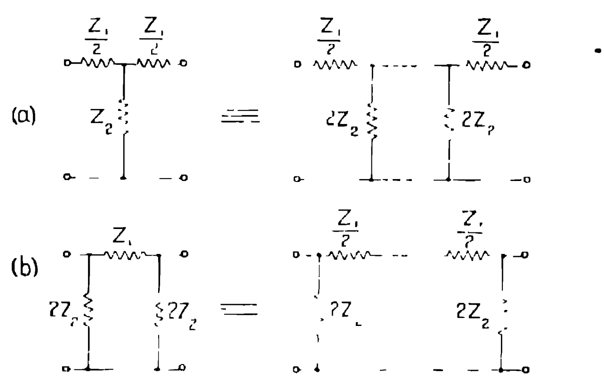


FIG. 672.—Prototype half-section as basic element of both T and π prototype sections.

Terminating half-sections

If an m-derived T section is split down the centre the result is an m-derived half-section as shown in Fig. 673*a*. Clearly one of the image impedances will be Z_{0T}, but the other one is *not* $Z_{0\pi}$; it may be shown to depend on the value of m. Let it be called $Z_{0\pi m}$. If an m-derived π section is divided down the centre, the result is a different kind of m-derived half-section, as in Fig. 673*b*. In this case, the image impedances are $Z_{0\pi}$ and Z_{0Tm}, where this latter impedance depends on the value of m.

These two types of m-derived half-sections, which are shown in

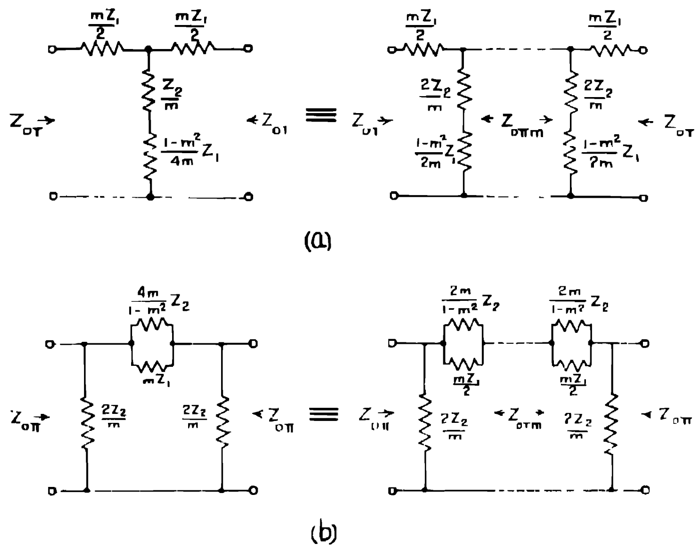


FIG. 673.—Derivation of m-derived half-sections.

Fig. 674, and the image impedances $Z_{0\pi m}$ and Z_{0Tm} are of great importance in filter theory; the reason will be apparent from Fig. 675, which shows how the values of the image impedances vary with frequency in the case of low-pass and high-pass m-derived half-sections.

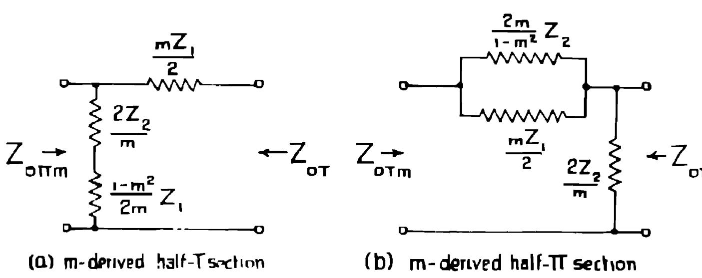


FIG. 674.—m-derived half sections showing image impedance.

The importance of "$m = 0{\cdot}6$ m-derived half-sections", or "terminating half-sections" as they are often called, lies in the fact that, when $m = 0{\cdot}6$, $Z_{0\pi m}$ and Z_{0Tm} lie between $0{\cdot}9\ R_0$ and $1{\cdot}1\ R_0$ over most of the pass-band. It will be seen that this enables a filter to be terminated accurately in a pure resistance (equal to the design impedance R_0) over practically the whole pass band.

Note that Z_{0Tm} and $Z_{0\pi m}$ are only the impedances obtained by splitting a T or π section in half. The impedance of a complete m-derived T or π section is the same as that of the prototype, *i.e.*, Z_{0T} or $Z_{0\pi}$.

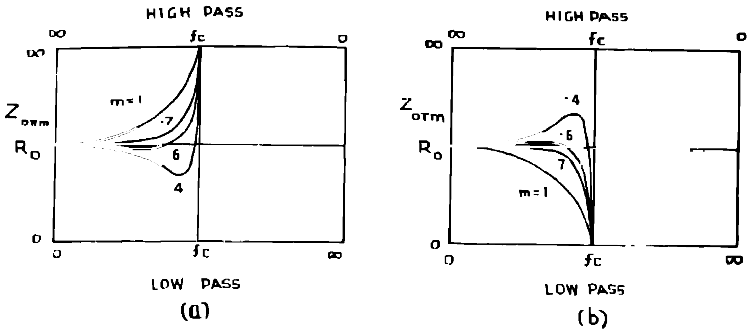


FIG. 675.—Variation of $Z_{0\pi m}$ and Z_{0Tm} with frequency for different values of m.

Example.—

Find the $m = 0{\cdot}6$ terminating half-sections required for a high-pass filter with $f_0 = 10$ kc/s and $R_0 = 600\Omega$.

The prototype section has already been found to have $L = 4.774$mH and $C = 0{\cdot}01326\mu$F (*see* example on page 649). By employing the method indicated by Figs. 661 and 662, the complete m-derived sections having $m = 0{\cdot}6$ may be determined, and are as shown in Fig. 676*a* and *b*. The m-derived T section divides symmetrically to give m-derived half-T sections as in Fig. 676*c*, and the corresponding m-derived half-π section is shown in Fig. 676*d*.

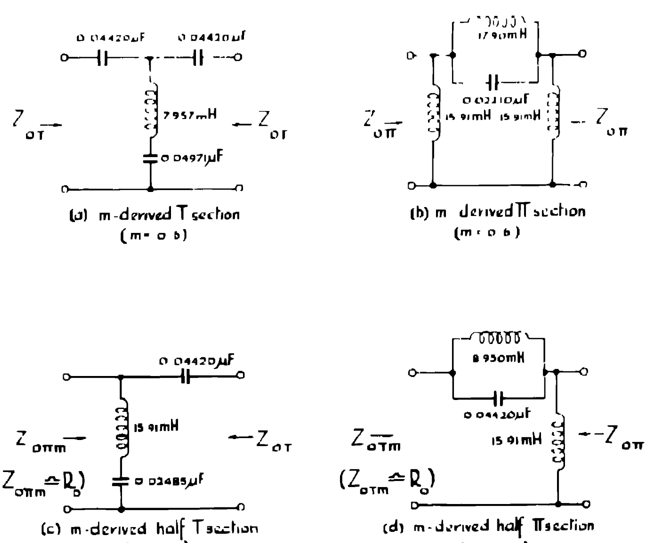
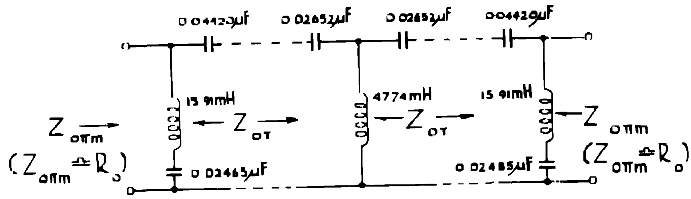
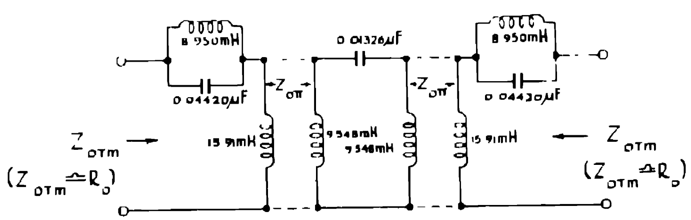


FIG. 676 *a–d*.



(e) m derived half-T sections used to terminate a prototype T section



(f) m derived half-π sections used to terminate a prototype π section

FIG. 676 e and f.—Illustrating use of m-derived half sections ($m = 0 \cdot 6$) for terminating filters.

Consider further how these sections may be utilised. Fig 676 shows how m-derived half-T sections are used to terminate a filter that is made up of T sections. In this particular case a single prototype T section has been terminated in this way. The impedance relationships shown in the diagram indicate how the terminating sections fit on to the prototype sections without mismatch, whilst presenting input and output impedances practically equal to the design impedance R_0 over the whole of the pass band. Fig. 676f shows, in a similar way, how m-derived half-π sections are used to terminate a filter that is made up of π sections.

Mismatch loss in the attenuation band

It has been seen that, by using suitable m-derived half-sections with $m = 0 \cdot 6$, it is possible to match the filter to purely resistive circuits, and this matching is practically perfect over about nine-tenths of the pass band. In the attenuation band, however, the characteristic impedance of the filter is a pure reactance, and the filter is terminated in a pure resistance (the design impedance R_0). The whole basis of argument so far has been the supposition made on page 636 that the filter is *always* correctly terminated in its characteristic impedance, that means, in a pure reactance, in the attenuation band.

The effect of this mis-termination may be deduced by treating the filter as a four-terminal network interposed between a generator and a load; the loss produced by the filter is then given by

equation 7 of Chapter 13 (*see* page 566). This method is somewhat tedious; however, it has been shown in Chapter 13 that when a network has a high attenuation, the input impedance is practically equal to Z_0, regardless of the termination.

This applies to all transmission networks, and, in particular, to filters. Since a filter has a high value of attenuation in the attenuation band, the fact that the termination is not equal to the characteristic impedance has little effect on the input impedance of the filter. This input impedance changes from resistance to reactance at the cut-off frequencies, and gives the pass and attenuation bands just as if the original supposition actually held good.

This mismatch in the attenuation band does, however, affect the overall attenuation of the section in that it gives rise to an additional "mismatch loss" over and above the calculated

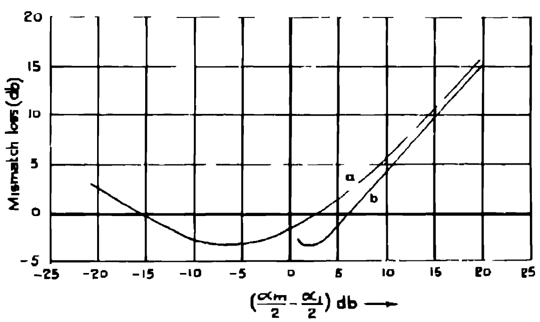


FIG. 677.— Mismatch loss per termination for an m-derived half section terminated in its design impedance (valid in the attenuation band provided $0{\cdot}3 < m < 0{\cdot}8$).

attenuation of the filter. The important case is that in which an m-derived half-section, used as a terminating section, is terminated in its design impedance. In this case it has been found that the mismatch loss is a function of the difference between the attenuation $\frac{\alpha_1}{2}$ of the prototype half-section and $\frac{\alpha_m}{2}$ that of the m-derived half-section. This applies for any frequency in the attenuation band, but only for sections with values of m between $0{\cdot}3$ and $0{\cdot}8$. The loss for *one* termination is given by the graphs in Fig. 677, curve b being used for frequencies on the pass band side of f_∞, and curve a for frequencies on the other side of f_∞.

Example.—

What is the mismatch loss at one termination of a low-pass filter at $1{\cdot}2f_0$ if an $m = 0{\cdot}6$ half-section is used for terminating?

The attenuation of an $m = 0{\cdot}6$ section at $1{\cdot}2f_0$ is (from

Fig. 665) $\alpha_m = 25$ db. The loss of the prototype at the same frequency is $\alpha_1 = 10$ db.

$$\therefore \frac{\alpha_m}{2} - \frac{\alpha_1}{2} = 7\cdot 5 \text{ db.}$$

This frequency is on the pass band side of f_∞ (for $f_\infty = 1\cdot 25\, f_c$) so curve b is used. This gives the mismatch loss as about 1 db.

Effect of resistance on filter characteristics

Reactances used in practical filters are not pure but contain resistance. In the case of a condenser this appears as a high shunt resistance and is usually negligible. In the case of inductance, it appears as series resistance and is often comparable in magnitude with the reactance. Its effect is to round off the attenuation characteristic near f_0 and f_∞ so that there will be attenuation in the pass band. Fig. 659 shows the characteristic of a typical

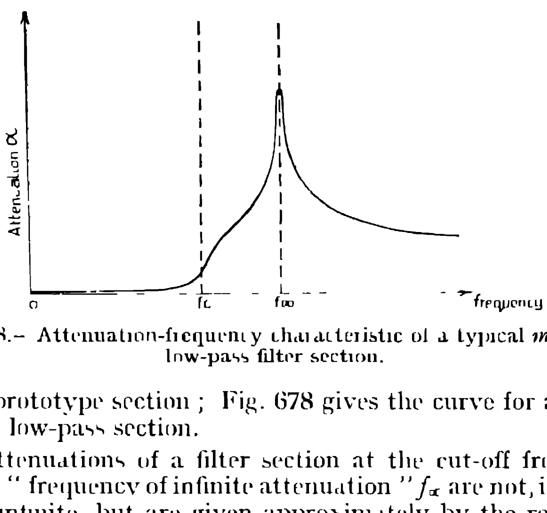


FIG. 678.—Attenuation-frequency characteristic of a typical m-derived low-pass filter section.

low-pass prototype section; Fig. 678 gives the curve for a practical m-derived low-pass section.

The attenuations of a filter section at the cut-off frequency f_c and at the "frequency of infinite attenuation" f_∞ are not, in practice, zero and infinite, but are given approximately by the relations:

$$\alpha_c = \frac{12}{m\sqrt{Q}} \quad \text{db} \tag{46a}$$

$$\alpha_\infty = 20\,.\log_{10}\frac{4m^2Q}{1-m^2} \quad \text{db} \tag{46b}$$

where $Q = \dfrac{\omega L}{R}$ for the inductance.

Complete filters

Complete filters are made up of a number of sections and half-sections connected together. The design of a complete filter is dependent upon two things: (i) the required attenuation characteristic, and (ii) the impedance. The first determines the cut-off frequency f_0; if this and the impedance are known, the elements of the prototype section can be calculated. The number of sections depends upon the attenuation required; if the attenuation

characteristic is to rise rapidly, at least one m-derived section with a small value of m will be required. $m = 0 \cdot 3$ to $0 \cdot 35$ is the value usually selected; a smaller value of m causes too large an attenuation at cut-off. If, in addition, the attenuation is to remain high at frequencies well beyond the cut-off frequency, either a prototype section or an m-derived section with a larger value of m (or both) will be required. Finally, if the impedance termination is important, two terminating half-sections with $m = 0 \cdot 6$ will be needed; it must be remembered that these two terminating half-sections will contribute to the total attenuation to the same extent as a complete section with $m = 0 \cdot 6$.

Example.—

Design a high-pass filter to satisfy the following conditions:—

(i) Attenuation above $10 \cdot 5$ kc/s to be less than 6 db.

(ii) Attenuation below $9 \cdot 5$ kc/s to be greater than 30 db.

(iii) Input and output impedances to be 600 ± 50 ohms above 12 kc/s.

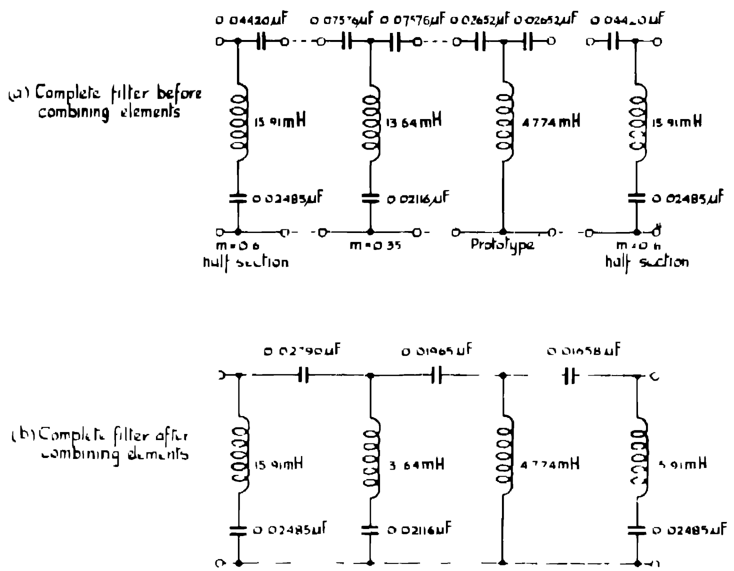


FIG. 679.

Suppose it is decided to work in terms of T sections throughout. This is in general an arbitrary decision, and in practice both types of filter would probably be worked out, the final selection being based on values and numbers of the components required in each case.

The cut-off frequency is selected as 10 kc/s, which fits in with conditions (i) and (ii). To satisfy condition (iii), two $m = 0 \cdot 6$

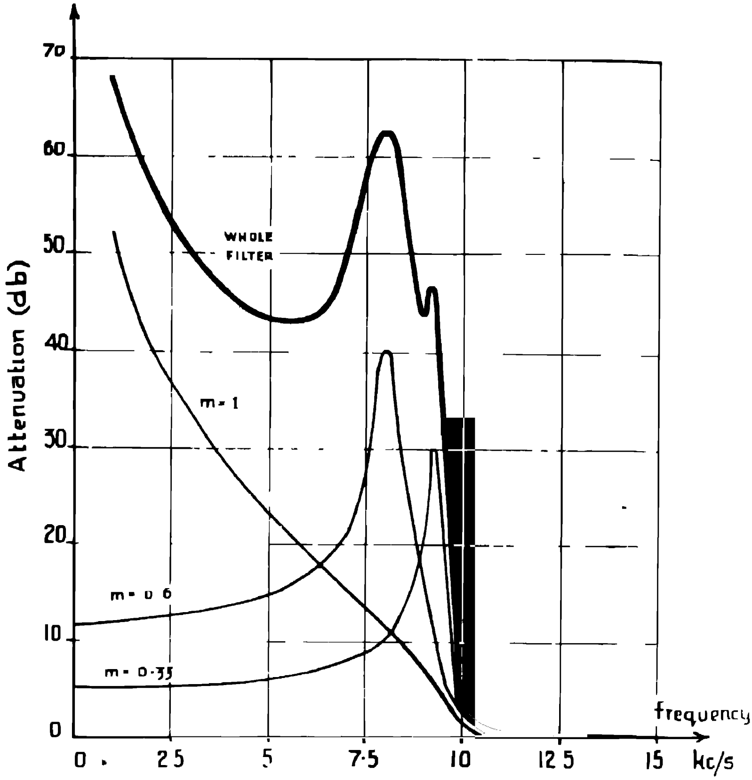


FIG 680 —Attenuation-frequency characteristics for complete high-pass filter and for component sections.

terminating half-sections are required. To satisfy condition (ii) at low frequencies, a prototype section is needed, and since a sharp cut-off is required an m-derived section with $m = 0 \cdot 35$ is included

All three of these sections have been determined in previous examples. The prototype section is shown in Fig. 658, the $m = 0 \cdot 35$ derived section in Fig 669a, and the terminating half-section in Fig. 676c. The complete filter is therefore as shown in Fig. 679a or, combining the series condensers, the final form is given in Fig. 679b. The attenuation-frequency curves for each section and for the whole filter are shown in Fig. 680.

BAND-PASS AND BAND-STOP FILTERS

Band-pass filters

There are several types of band-pass filter, of which one only will be discussed here. This is the constant-k type, and a prototype T section is shown in Fig. 681. It will be noted that the series

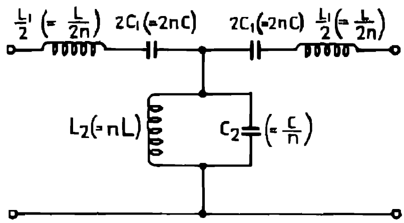


FIG 681 —Prototype band-pass filter section.

and shunt arms are arranged to have the same resonant frequency; let this be called f_0. It will be seen that :—

$$f_0 = \frac{1}{2\pi\sqrt{LC}} \quad i.e. \ \omega_0 = \frac{1}{\sqrt{LC}} \tag{47}$$

To show that the filter is constant-k, *i.e.* that the product of the series and shunt arm impedances is equal to a real constant, put :—

$$\sqrt{\frac{L}{C}} = R_0 \tag{48}$$

and let
$$x = \frac{f}{f_0} = \frac{\omega}{\omega_0} = \omega\sqrt{LC} \tag{49}$$

so that
$$\omega L = R_0 x \tag{50}$$

and
$$\frac{1}{\omega C} = \frac{R_0}{x} \tag{51}$$

Now the series arm $Z_1 = j\frac{\omega L}{n} - \frac{j}{n\omega C} = j\frac{R_0 x}{n} - \frac{jR_0}{nx}$

i.e.
$$Z_1 = j\frac{R_0(x^2 - 1)}{nx} \tag{52}$$

and the shunt arm $Z_2 = \dfrac{j\omega nL\dfrac{-jn}{\omega C}}{j\omega nL - \dfrac{jn}{\omega C}} = \dfrac{-jn\dfrac{L}{C}}{\omega L - \dfrac{1}{\omega C}} = \dfrac{jnR_0^2}{\dfrac{R_0}{x} - R_0 x}$

i.e.
$$Z_2 = j\frac{nR_0 x}{1 - x^2} \tag{53}$$

Multiplying equations 52 and 53, it is seen that :—

$$Z_1 Z_2 = R_0^2$$

and the section is constant-k; equation 48 therefore gives the design impedance.

To verify that such a section has a band-pass characteristic, let $Z_1 = jX_1$ and $Z_2 = jX_2$, and draw the reactance-frequency sketches for X_1 and $\left(\frac{X_1}{4} + X_2\right)$, as in Fig. 682.

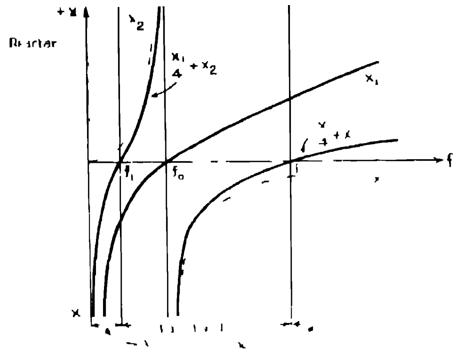


FIG. 682 — Reactance-frequency sketch for prototype band-pass filter section.

It is apparent that between f_1 and f_2 (the two series resonant frequencies of $\frac{Z_1}{4} + Z_2$) the reactance curves are on opposite sides of the frequency axis and the filter has a pass band; outside these limits of frequency, the curves are on the same side of the frequency axis and the filter attenuates. The network is therefore a band pass filter with cut-off frequencies f_1 and f_2. It would be possible to find f_1 and f_2 by evaluating the series resonant frequencies of $\frac{Z_1}{4} + Z_2$, but a simpler method will be employed.

Consider the characteristic impedance of the section —

$$Z_0 = \sqrt{\frac{Z_1^2}{4} + Z_1 Z_2}$$

Then, from equations 52 and 53 :—

$$Z_0 = \sqrt{R_0^2 - \frac{R_0^2(x^2-1)^2}{4n^2x^2}} = R_0\sqrt{1 - \frac{(x^2-1)^2}{4n^2x^2}}$$

Thus Z_0 is real, giving a *pass band* if : —

$$\frac{(x^2-1)^2}{4n^2x^2} < 1$$

i.e.

$$-1 < \frac{x^2-1}{2nx} < 1$$

The cut-off frequencies are given by the positive roots of the equations :—

$$\frac{x^2-1}{2nx} = -1 \quad \text{and} \quad \frac{x^2-1}{2nx} = +1$$

Hence :—

$$x^2 + 2nx - 1 = 0 \qquad\qquad x^2 - 2nx - 1 = 0$$

$$\therefore\ x^2 + 2nx + n^2 = 1 + n^2 \qquad\qquad \therefore\ x^2 - 2nx + n^2 = 1 + n^2$$

$$\therefore\ x + n = \pm\sqrt{1 + n^2} \qquad\qquad \therefore\ x_2 - n = \pm\sqrt{1 + n^2}$$

$$\therefore\ x_1 = -n + \sqrt{1 + n^2} \quad (54) \qquad\qquad \therefore\ x_2 = +n + \sqrt{1 + n^2} \quad (55)$$

These results give the cut-off frequencies, since $x_1 = \frac{f_1}{f_0}$ and $x_2 = \frac{f_2}{f_0}$

Note that $x_2 - x_1 = 2n$

i.e. $f_2 - f_1 = 2nf_0$ (i) (56)

an expression giving the *band-width* in terms of n and f_0

Also from equations 54 and 55

$$x_1 x_2 = 1 \quad \text{(ii)}$$

$$f_1 f_2 = f_0^2 \qquad (57)$$

i.e. f_0 is the mid-band frequency, where mid-band frequency is taken to be the geometric mean of the two cut-off frequencies.

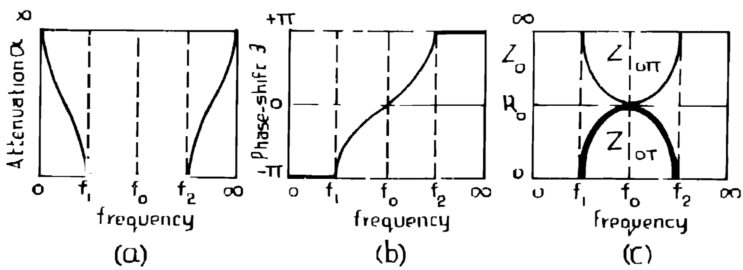


Fig. 683. Prototype band-pass filter section.—Variation with frequency of

(a) Attenuation (b) Phase-shift (c) Characteristic impedance

It can be shown that the attenuation outside the pass band is given by

$$\alpha = \cosh^{-1}\left\{\frac{(1 - x^2)^2}{2n^2x^2} - 1\right\} \text{ nepers} \qquad (58)$$

This gives the characteristic of Fig. 683*a*. The phase shift is given by :—

$$\beta = \cos^{-1}\left\{1 - \frac{(1 - x^2)^2}{2n^2x^2}\right\} \text{ radians} \qquad (59)$$

giving the curve of Fig. 683*b*.

$$Z_{0T} = R_0\sqrt{1 - \frac{(x^2 - 1)^2}{4n^2x^2}} \qquad (60)$$

and :—

$$Z_{0\pi} = \frac{R_0}{\sqrt{1 - \frac{(x^2 - 1)^2}{4n^2 x^2}}} \qquad (61)$$

plotted in Fig. 683c.

Design of a prototype band-pass filter

Suppose that the required cut-off frequencies f_1 and f_2 and the design impedance R_0 are given ; equation 57 then gives :—

$$f_0 = \sqrt{f_1 f_2}$$

and 56 gives :—

$$n = \frac{f_2 - f_1}{2f_0}$$

Thus f_0 and n may be found.

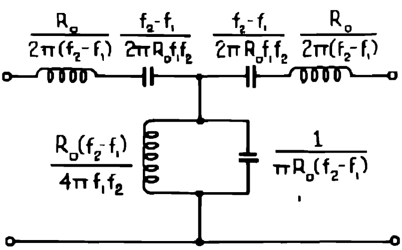


FIG 684.—Prototype band-pass filter section with component values in terms of cut-off frequencies and design impedance

Then, from equations 47 and 48 :—

$$L = \frac{R_0}{2\pi f_0} \qquad (62)$$

$$C = \frac{1}{2\pi f_0 R_0} \qquad (63)$$

Hence
$$L_1 = \frac{L}{n} = \frac{R_0}{\pi(f_2 - f_1)} \qquad (64)$$

$$L_2 = nL = \frac{R_0(f_2 - f_1)}{4\pi f_1 f_2} \qquad (65)$$

$$C_1 = nC = \frac{f_2 - f_1}{4\pi R_0 f_1 f_2} \qquad (66)$$

$$C_2 = \frac{C}{n} = \frac{1}{\pi R_0 (f_2 - f_1)} \qquad (67)$$

The required section is therefore as shown in Fig. 684.

Note that the inductance in the series arm and the capacity in the shunt arm depend only on the band-width and the design

impedance, and not on the position of the band. This is worth noting in connection with multi-channel VF telegraph systems, where all the filters have the same band-width.

Example.—

Design a prototype T section for a band-pass filter having cut-off frequencies of 1000 c/s and 4000 c/s, and a design impedance of 600 Ω. In this case $R_0 = 600, f_1 = 1000$, and $f_2 = 4000$. Putting these values into Fig. 684 : —

$$\text{Inductance in series arm is } \frac{600}{2\pi\,(4000-1000)} = \underline{31 \cdot 83\text{mH}}$$

$$\text{Capacity in series arm is } \frac{4000-1000}{2\pi 600 \, . \, 1000 \, . \, 4000} = 0 \cdot 1989\mu\text{F}$$

$$\text{Inductance in shunt arm is } \frac{600\,(4000-1000)}{4\pi\, 1000 \, . \, 4000} = \underline{35 \cdot 80\text{mH}}$$

$$\text{Capacity in shunt arm is } \frac{1}{\pi \, . \, 600\,(4000-1000)} = 0 \cdot 1768\mu\text{F}$$

The required section is therefore as shown in Fig. 685.

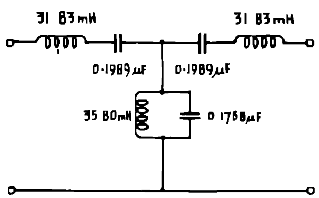


FIG 685.—Illustrating design procedure for band-pass prototype T section.

m-derived band-pass sections

Prototype band-pass filters may be derived in a similar manner to low- and high-pass filters. From Figs. 660 and 681 the m-derived

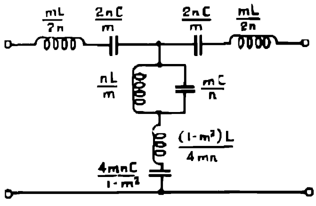


FIG. 686.—m-derived T section for a band-pass filter.

section of the band-pass filter is seen to be as shown in Fig. 686. The frequency or frequencies of infinite attenuation can be calculated from the series resonance of the shunt arm.

The impedance of the shunt arm is given by :—

$$Z'_2 = \frac{Z_2}{m} + \frac{Z_1(1-m^2)}{4m}$$ Pag 651 eq No

where Z_1 and Z_2 are the series and shunt impedances respectively of the prototype

From equations 52 and 53

$$Z_1 = j\frac{R_0(x^2-1)}{nx} \quad \text{and} \quad Z_2 = -j\frac{R_0 nx}{(x^2-1)}$$

$$\therefore \qquad Z'_2 = \frac{-jR_0 nx}{m(x^2-1)} + \frac{jR_0(x^2-1)(1-m^2)}{4mnx}$$

$$= \frac{jR_0[(1-m^2)(x^2-1)^2 - 4n^2x^2]}{4mnx(x^2-1)}$$

Therefore the series resonant condition is

$$(x^2-1)^2 = \frac{4n^2x^2}{1-m^2}$$

$$\therefore \qquad x^2 - 1 = \pm\frac{2nx}{\sqrt{1-m^2}}$$

$$\therefore \quad x^2 - \frac{2nx}{\sqrt{1-m^2}} - 1 = 0 \quad \text{or} \quad x^2 + \frac{2nx}{\sqrt{1-m^2}} - 1 = 0$$

$$\therefore \quad x = \frac{n}{\sqrt{1-m^2}} \pm \sqrt{\frac{n^2}{1-m^2} + 1} \quad \text{or}$$

$$x = -\frac{n}{\sqrt{1-m^2}} \pm \sqrt{\frac{n^2}{1-m^2} + 1}$$

This gives four values of x, but only two of them are positive, calling these $x_{1\infty}$ and $x_{2\infty}$, then

$$x_{1\infty} = \sqrt{1 + \frac{n^2}{1-m^2}} - \frac{n}{\sqrt{1-m^2}} \tag{68}$$

$$x_{2\infty} = \sqrt{1 + \frac{n^2}{1-m^2}} + \frac{n}{\sqrt{1-m^2}} \tag{69}$$

Note that $\quad x_{1\infty} \cdot x_{2\infty} = 1 \qquad (70)$

and that $\quad x_{2\infty} - x_{1\infty} = \dfrac{2n}{\sqrt{1-m^2}} \qquad (71)$

• But $x = \dfrac{f}{f_0}$, so that $x_{1\infty}$ (a value of x giving infinite attenuation), is $\dfrac{f_{1\infty}}{f_0}$, where $f_{1\infty}$ is a frequency of infinite attenuation. Similarly $x_{2\infty} = \dfrac{f_{2\infty}}{f_0}$, where $f_{2\infty}$ is the other frequency of infinite attenuation.

From equation 70 :—

$$f_{1\infty} \cdot f_{2\infty} = f_0^2 \tag{72}$$

so that f_0 is the geometric mean of the frequencies of infinite attenuation as well as being the geometric mean of the cut-off frequencies

From equations 71 and 56 —

$$f_{2\infty} - f_{1\infty} = \frac{2\pi f_0}{\sqrt{1 - m^2}} = \frac{\text{Bandwidth}}{\sqrt{1 - m^2}} \tag{73}$$

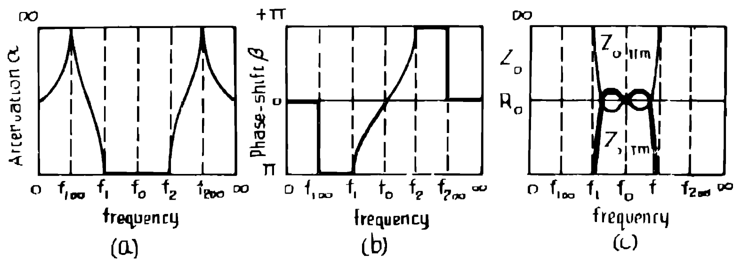


FIG. 687 *m* derived band pass filter sections Variation with frequency of
(a) Attenuation (b) Phase shift (c) $Z_{0\pi m}$ and Z_{0Tm}

Fig 687*a* and *b* shows the attenuation frequency and phase-shift frequency characteristics for *m* derived band pass sections Fig 687*c* shows $Z_{0\pi m}$ and Z_{0Tm} plotted against frequency $Z_{0\pi}$ and Z_{0T} are the same as for the prototype section

It has been shown that the shunt arm of the *m*-derived band-pass filter has two series resonant frequencies, being the frequencies of infinite attenuation It can be proved that this shunt arm may be replaced by two series resonant circuits in parallel (*see* Fig 688)

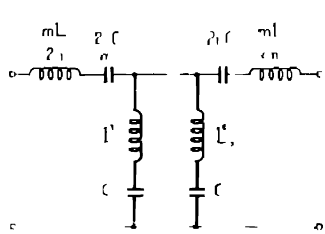


FIG 688 —Alternative form of *m* derived band pass section

The corresponding values of L'_1, C'_1, and L'_2, C'_2 are :

$$L'_1 = \frac{L}{n}\,\frac{(1 - m^2)}{4m} \cdot \frac{x_{1\infty} + x_{2\infty}}{x_{1\infty}} \cdot \frac{n}{m} \cdot \frac{1 + x_{1\infty}^2}{(1 - x_{1\infty}^2)^2}\, I \tag{74}$$

$$C_1' = Cn\frac{4m}{1-m^2}\cdot\frac{1}{x_{1\infty}(x_{1\infty}+x_{2\infty})} = \frac{m}{n}\,\frac{(1-x_{1\infty}^2)^2}{x_{1\infty}^2(1+x_{1\infty}^2)}\,C \quad (7\text{-}$$

$$L_2' = \frac{L}{n}\cdot\frac{1-m^2}{4m}\cdot\frac{x_{1\infty}+x_{2\infty}}{x_{2\infty}} = \frac{n}{m}\,\frac{1+x_{2\infty}^2}{(1-x_{2\infty}^2)^2}\,L \quad (76$$

$$C_2' = Cn\frac{4m}{1-m^2}\cdot\frac{1}{x_{2\infty}(x_{1\infty}+x_{2\infty})} = \frac{m}{n}\,\frac{(1-x_{2\infty}^2)^2}{x_{2\infty}^2(1+x_{2\infty}^2)}\,C \quad (77$$

It will be noticed that the resonant frequencies of the two series resonant circuits are $f_{1\infty}$ and $f_{2\infty}$ respectively. This form of the *m*-derived section is the one that is almost invariably encountered in practice (*see* Fig. 689).

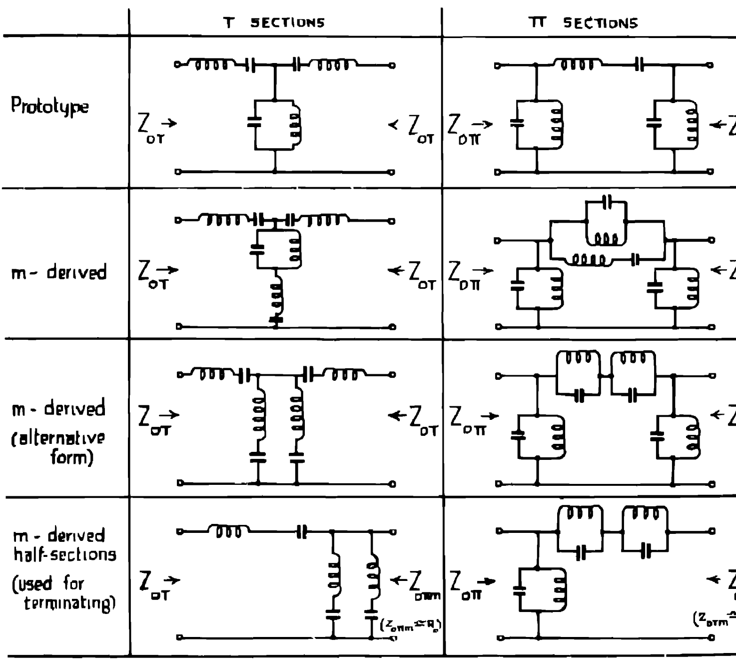


FIG. 689 —Summary of band-pass filter sections

Band-stop filters

Fig. 690 shows the prototype T section of a band-stop filter. As in the case of the band-pass filter, the series and shunt arms are arranged to have the same resonant frequency.

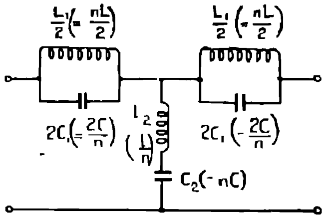


FIG. 690.—Prototype T section band-stop filter

Let this frequency be f_0. It will be seen that :—

$$f_0 = \frac{1}{2\pi\sqrt{LC}}, \text{ or } \omega_0 = \frac{1}{\sqrt{LC}} \tag{78}$$

To show that the filter is constant-k,

put $$\sqrt{\frac{L}{C}} = R_0 \tag{79}$$

and let $$x = \frac{f}{f_0} = \frac{\omega}{\omega_0} = \omega\sqrt{LC} \tag{80}$$

so that $$\omega L = R_0 x \tag{81}$$

and $$\frac{1}{\omega C} = \frac{R_0}{x} \tag{82}$$

Now in this case,

$$\text{the series arm } Z_1 = \frac{j\omega nL \times \frac{-jn}{\omega C}}{j\omega nL - \frac{jn}{\omega C}}$$

$$= \frac{-jn\frac{L}{C}}{\omega L - \frac{1}{\omega C}}$$

$$= \frac{jnR_0^2}{\frac{R_0}{x} - R_0 x}$$

$$= j\frac{nR_0 x}{1 - x^2} \tag{83}$$

$$\text{and the shunt arm } Z_2 = \frac{j\omega L}{n} - \frac{j}{n\omega C}$$

$$= j\frac{R_0 x}{n} - \frac{jR_0}{nx}$$

$$= j\frac{R_0(x^2 - 1)}{nx} \tag{84}$$

Multiplying equations 83 and 84, it is seen that

$$Z_1Z_2 = R_0^2$$

and the section is a constant-k section, equation 79 therefore g
the design impedance

Fig. 691 gives the reactance-frequency sketch for X_1 and $\frac{X_1}{4}$
and shows that the section is a band-stop filter with cut-o
frequencies f_1 and f_2, which are the series resonant frequen
of $\left(\frac{Z_1}{4} + Z_2\right)$.

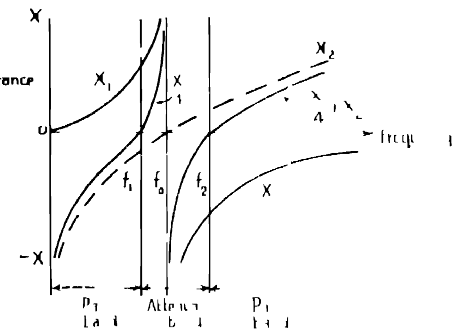


FIG. 691 — Reactance frequency sketch for prototype band stop filter

Consider Z_0 for the section

$$Z_0 = \sqrt{\frac{Z_1^2}{4} + Z_1Z_2}$$

Then, from equations 83 and 84 —

$$Z_0 = \sqrt{R_0^2 - \frac{n^2R_0^2x^2}{4(1-x^2)^2}} = R_0\sqrt{1 - \frac{n^2x^2}{4(1-x^2)^2}}$$

It can be shown that the cut off frequencies are given by

$$x_1 = \frac{-n + \sqrt{n^2+16}}{4} \quad \text{and} \quad x_2 = \frac{n + \sqrt{n^2+16}}{4} \tag{85}$$

$$\text{or} \quad f_1 = f_0\,\frac{-n + \sqrt{n^2+16}}{4} \quad \text{and} \quad f_2 = f_0\,\frac{n + \sqrt{n^2+16}}{4} \tag{86}$$

Note that $$x_2 - x_1 = \frac{n}{2}$$

i.e. $$f_2 - f_1 = \frac{nf_0}{2} \tag{87}$$

giving the *bandwidth* in terms of n and f_0

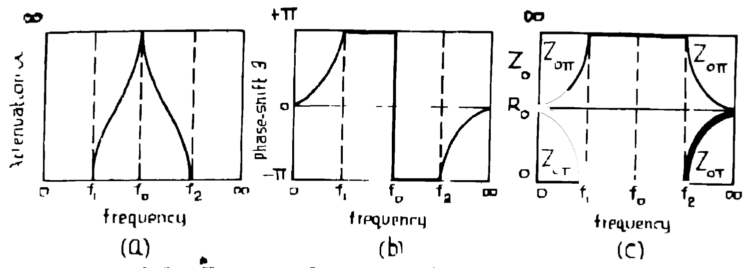


FIG. 692.—Prototype band-stop filter sections—Variation with frequency of
(a) Attenuation. (b) Phase-shift. (c) Characteristic impedance.

Also, that $x_1 x_2 = 1$

i.e. $$f_1 f_2 = f_0^2 \tag{88}$$

i.e., f_0, the *mid-band frequency*, is the geometric mean of the two cut-off frequencies. Fig 692 shows the variation with frequency of attenuation, phase-shift and characteristic impedance.

Design of a prototype band-stop filter

Let the required cut-off frequencies be f_1 and f_2 and the design impedance R_0.

Equation 88 then gives:

$$f_0 = \sqrt{f_1 f_2}$$

and equation 87 gives:

$$n = \frac{2(f_2 - f_1)}{f_0}$$

Thus f_0 and n may be found.

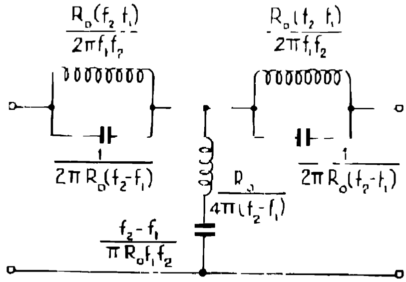


FIG. 693.—Prototype band-stop filter section with component values in terms of cut-off frequencies and design impedance.

From equations 78 and 79

$$L = \frac{R_0}{2\pi f_0} \quad \text{and} \quad C = \frac{1}{2\pi f_0 R_0}$$

Hence
$$L_1 = nL = \frac{R_0(f_2 - f_1)}{\pi f_1 f_2} \tag{89}$$

$$L_2 = \frac{L}{n} = \frac{R_0}{4\pi(f_2 - f_1)} \tag{90}$$

$$C_1 = \frac{C}{n} = \frac{1}{4\pi R_0(f_2 - f_1)} \tag{91}$$

$$C_2 = nC = \frac{f_2 - f_1}{\pi R_0 f_1 f_2} \tag{92}$$

The required section is therefore as shown in Fig. 693.

It is interesting to note that if a prototype band-pass half-section is taken and the series and shunt arms are interchanged, a band-stop half-section is formed that has the same cut-off frequencies as the original band-pass half-section.

m-derived band-stop section

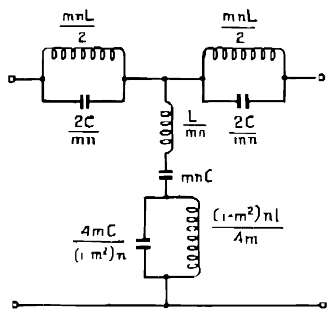


FIG. 694.—*m*-derived band-stop filter section.

m-derived band-stop sections can be obtained in the same way as for band-pass sections; Fig. 694 shows an *m*-derived band-pass T section. As with *m*-derived band-pass sections, the alternative

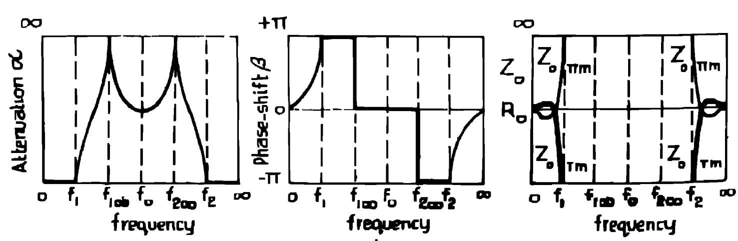


FIG. 695.—*m*-derived band-stop filter sections—Variation with frequency of

(*a*) Attenuation. (*b*) Phase-shift. (*c*) $Z_{0\pi m}$ and Z_{0Tm}.

form of the shunt arm is almost invariably used (*see* Fig. 696). Fig. 695*a* and *b* shows the attenuation-frequency and phase-shift frequency characteristics for an *m*-derived band-stop filter. Fig. 695*c* shows $Z_{0\pi m}$ and Z_{0Tm} plotted against frequency. $Z_{0\pi}$ and Z_{0T} are the same as for the prototype section.

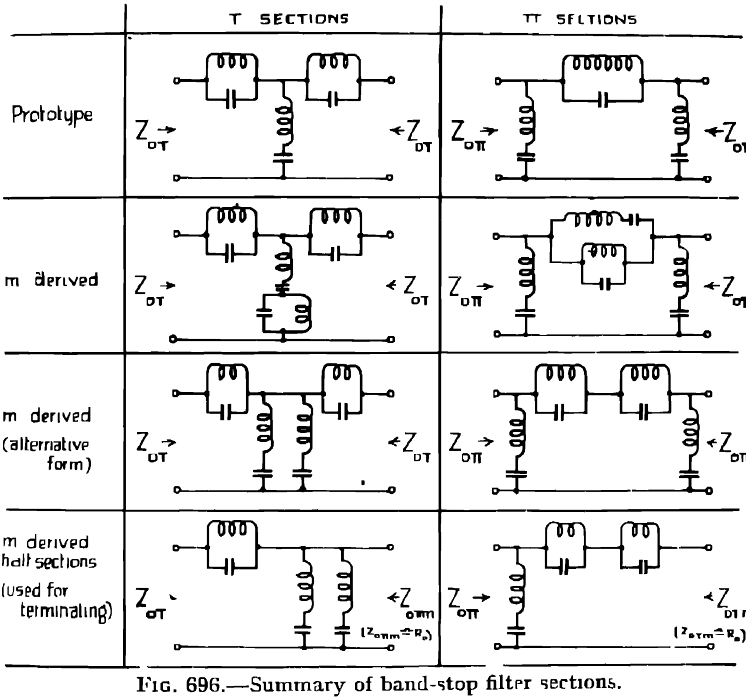


FIG. 696.—Summary of band-stop filter sections.

FURTHER FILTER DESIGN CONSIDERATIONS

Connection of filters in parallel

Consider the case of a high-pass and a low-pass filter connected in parallel at one end, the cut-off frequencies being almost equal (*see* Fig. 697*a*).

In the pass band of the low-pass filter, the high-pass filter adds a reactance in parallel with the resistive termination of the low-pass filter. This reactance can be represented approximately by a condenser and inductance in series, as in Fig. 697*b*. The low-pass filter is thus incorrectly terminated in its pass band

If, however, a series inductance be added to the low-pass filter as in Fig. 697c (i.e. the final series inductance of the low-pass filter be increased), the added series inductance, together with the shunting reactance of the high pass filter will form an approximate m-derived terminating half-section for the low-pass filter.

For an $m = 0.6$ termination the added inductance will be 0.6 of the series inductance of a prototype half-section. Similarly the

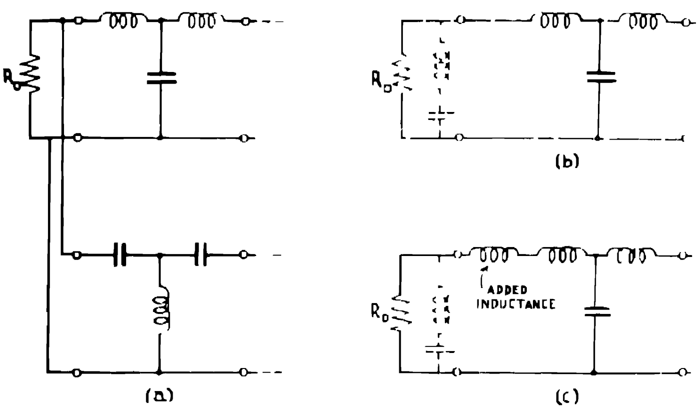


Fig. 697. High pass and low pass filters in parallel.
(a) High pass and low pass filters in parallel.
(b) Equivalent circuit in the low pass range.
(c) Added inductance giving half section termination.

final condenser of the high pass filter may be modified so that in the high-pass range the high pass filter is terminated in an equivalent m-derived half-section ($m = 0.6$). If both these modifications are carried out, both filters will be correctly terminated over their pass ranges.

Band-pass filters

The same principle applies when a number of band pass filters with adjacent pass bands are connected in parallel. For considering any one filter, the shunt reactance due to all the other filters in

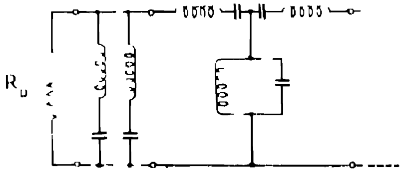


Fig. 698.—Equivalent circuit in the pass band of one of several band pass filters in parallel.

parallel can be represented (at frequencies just outside the pass band) by two series resonant circuits in parallel (*see* Fig. 698).

The resonant frequencies of these circuits will be respectively just above and just below the cut-off frequencies of the filter under consideration. These resonant frequencies may be regarded as f_1, and f_2, for the filter considered; for these circuits are, in fact, the shunt arm of an *m*-derived band-pass filter (*see* Fig. 688). Thus to give an $m = 0 \cdot 6$ half-section termination, it is necessary to increase the last series impedance of each band-pass filter by $0 \cdot 6$ of the series impedance of the prototype half-section.

Compensating networks

The above argument applies only to a band-pass filter that has at least one other filter above and below it; it does not hold for the first and last of a number of adjacent filters. To provide correct terminations for these filters, some additional shunt network must be provided. Such a network is called a "compensating network", and consists of two series resonant circuits in parallel (or any circuit with two series resonances); these resonant frequencies are arranged just below the lower cut-off frequency of the first filter and just above the higher cut-off frequency of the last filter. These resonant circuits take the place of the two "missing" band-pass filters, and thus complete the correct *m*-derived terminating half-sections for the first and last filters.

Impedance transformation in band-pass filters

It is often useful to be able to transform the impedance inside a band-pass filter particularly when component values would otherwise be inconveniently large or small. The use of ordinary transformers for this purpose would involve taking into account such considerations as the leakage inductance, the primary inductance,

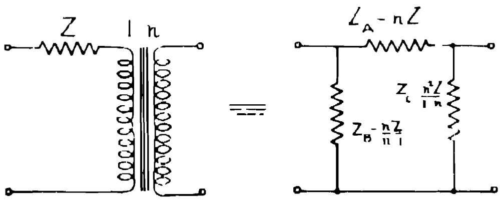


FIG. 699.—π section equivalent of an ideal transformer and series impedance.

and the capacity of the transformer windings, all of which would affect the performance of the filter. Fortunately if Z represents a capacity (*see* Fig. 699), it is possible to find an equivalent circuit for the transformer in terms of capacities only.

Fig. 699 shows an ideal transformer of turns ratio $1 : n$ with a series impedance Z in the primary, together with its equivalent π section which was given in Chapter 13. This equivalence may be

verified by equating open- and short-circuit input impedances, as follows :—

Let Z_1 be the impedance looking into the primary with the secondary short-circuited.

Then $$Z_1 = Z = \frac{Z_A Z_B}{Z_A + Z_B} \tag{93}$$

Let Z_2 be the impedance looking into the secondary with the primary short-circuited.

Then $$Z_2 = n^2 Z = \frac{Z_A Z_C}{Z_A + Z_C} \tag{94}$$

The impedance looking into either winding with the other open-circuited is infinite (since it is an ideal transformer).

$$\therefore \quad \frac{Z_B (Z_A + Z_C)}{Z_A + Z_B + Z_C} = \frac{Z_C (Z_A + Z_B)}{Z_A + Z_B + Z_C} = \infty$$

Hence $$Z_A + Z_B + Z_C = 0 \tag{95}$$

From (93) and (95) :—

$$Z = -\frac{Z_A Z_B}{Z_C} \tag{96}$$

From (94) and (95) :—

$$n^2 Z = -\frac{Z_A Z_C}{Z_B} \tag{97}$$

Multiplying (96) and (97) :—

$$n^2 Z^2 = Z_A^2$$

$$\therefore \quad Z_A = nZ \tag{98}$$

(The other root, $Z_A = -nZ$, actually gives rise to a second equivalent circuit.)

From (95) and (98) :—

$$Z_B + Z_C = -nZ \tag{99}$$

From (96) and (98) :—

$$\frac{Z_B}{Z_C} = -\frac{1}{n} \tag{100}$$

Eliminating Z_B from (99) and (100) :—

$$Z_C - \frac{Z_C}{n} = -nZ$$

$$\therefore \quad Z_C = \frac{n^2 Z}{1 - n} \tag{101}$$

From (100) and (101) :—

$$Z_B = \frac{nZ}{n - 1} \tag{102}$$

Thus an equivalent π network is as shown in Fig. 699.

The use of this transformation will be illustrated by an example of band-pass filter design taken from a multi-channel VF telegraph system.

Design of band-pass filters for a multi-channel VF telegraph system

All the band-pass channel filters have a design impedance $R_0 = 600\Omega$, a bandwidth $f_2 - f_1 = 120$ c/s, and mid-band frequencies 420, 540, 660 and so on up to 1980 c/s. From equations

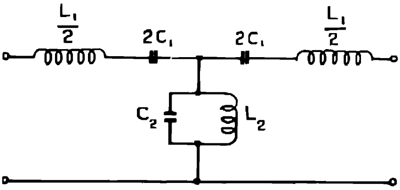


FIG. 700.—Prototype band-pass T section.

64 to 67, the values for a prototype, as shown in Fig. 700, are :—

$$L_1 = \frac{R_0}{\pi (f_2 - f_1)},\ C_1 = \frac{f_2 - f_1}{4\pi R_0 f_1 f_2},\ L_2 = \frac{R_0 (f_2 - f_1)}{4\pi f_1 f_2},\ C_2 = \frac{1}{\pi R_0 (f_2 - f_1)}$$

Hence all the channel filters have :—

$$\frac{L_1}{2} = \frac{600}{2\pi \,.\, 120} = 0{\cdot}797\text{H and } C_2 = \frac{1}{\pi \,.\, 600 \,.\, 120} = 4{\cdot}421\mu\text{F}$$

Taking as an example the filter with a mid-band frequency of 1500 c/s :—

$$2C_1 = \frac{2 \,.\, 120 \,.\, 10^6}{4\pi \,.\, 600 \,.\, 1500^2} = 0{\cdot}01415\mu\text{F}$$

and
$$L_2 = \frac{600 \,.\, 120}{4\pi \,.\, 1500^2} = 2{\cdot}55\text{mH}$$

Note that the value of C_2 ($4{\cdot}421\mu$F) is large and the inductance L_2 is small. C_2 would have to be a paper dielectric condenser, owing to the large cost and bulk of a mica condenser of this capacity; but paper condensers are not stable enough for filter construction, and tend to have a high loss. L_2 would have a very small number of turns (say about 100) and an inaccuracy of ± 1 turn might affect the inductance by as much as ± 2 per cent., which is not sufficiently accurate for this filter. If the impedance at the middle of the section could be increased, a larger inductance and a smaller capacity could be used, having of course the same resonant frequency as the original shunt circuit. This can be done in theory by the use of two ideal transformers, as shown in Fig. 701*a*.

Alternatively it can be done by replacing the ideal transformer and the series condenser $2C_1$ by its equivalent π section as in Fig. 701*b*. The final form of the filter is as shown in Fig. 701*c*, the various shunt capacities having been combined where possible.

n can have any value that makes the components of the final section physically realisable. For example, n must be greater than one, or the capacities $\frac{2(n-1)}{n} C_1$ will not be realisable. A

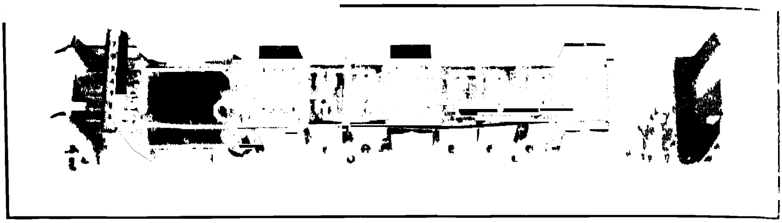


PLATE 34 —Multi-channel VF telegraph band pass filter, mid-band frequency 1500 c/s

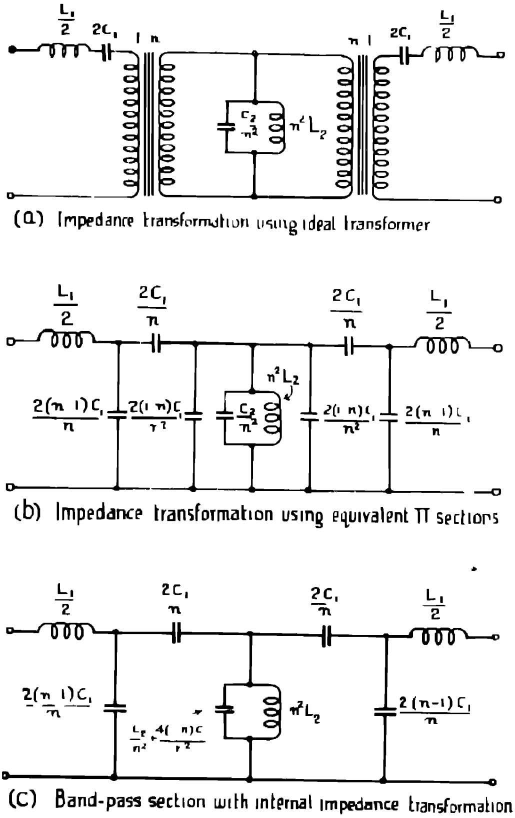


FIG 701 —Prototype band-pass T section with internal impedance transformation.

convenient value of n is that value which makes all the shunt capacities equal. This gives an equation :—

$$\frac{2(n-1)C_1}{n} = \frac{C_2}{n^2} + \frac{4(1-n)}{n^2}C_1$$

$$2n(n-1)C_1 = C_2 + 4(1-n)C_1$$

$$2n^2 + 2n = \frac{C_2}{C_1} + 4$$

$$4n^2 + 4n + 1 = \frac{2C_2}{C_1} + 9$$

$$n = \tfrac{1}{2}\left[\sqrt{\frac{2C_2}{C_1} + 9} - 1\right] \qquad (103)$$

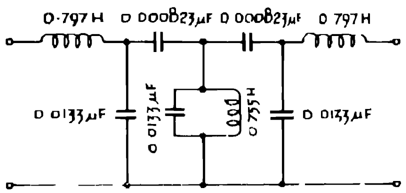


FIG. 702. - Illustrating design of prototype band-pass filter with internal impedance transformation.

In the case under consideration :—

$$C_2 = 4\cdot421\mu\text{F} \quad \text{and} \quad C_1 = 0\cdot007075\mu\text{F}$$

$$\therefore \qquad n = \tfrac{1}{2}\left[\sqrt{\frac{8\cdot842}{0\cdot007075} + 9} - 1\right] = \tfrac{1}{2}\left[35\cdot4 - 1\right]$$

$$\therefore \qquad n = 17\cdot2 \quad \text{and} \quad n^2 = 296$$

The shunt condensers now become : -

$$\frac{2(n-1)}{n}C_1 = 0\cdot0133\mu\text{F}$$

The series condensers are then : -

$$\frac{2C_1}{n} = 0\cdot000823\mu\text{F}$$

and the shunt inductance is :—

$$n^2L_2 = 0\cdot755\ \text{H}$$

Hence the complete section is as shown in Fig. 702

If, as is the case in a multi-channel VF telegraph system, the filters are all connected in parallel at one end, the series impedances

at one end will have to be increased by 0·6 of their value (to give an $m = 0{\cdot}6$ termination). This must be done before the transformation of the central section is carried out; that is, the prototype is first modified, and then the transformation is carried out using the same value of n as before.

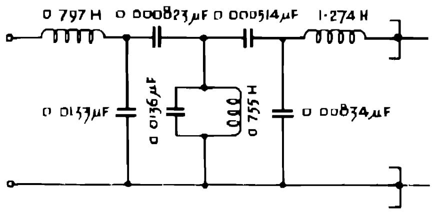


FIG. 703.—Prototype band-pass filter with internal impedance transformation, modified for parallel connection with other filters.

Fig. 703 shows the complete filter after modification in this way The other channel filters are connected in parallel with the right hand end.

LATTICE FILTER SECTIONS

The majority of filters in common use are of the ladder type, but "lattice" filters are occasionally encountered.

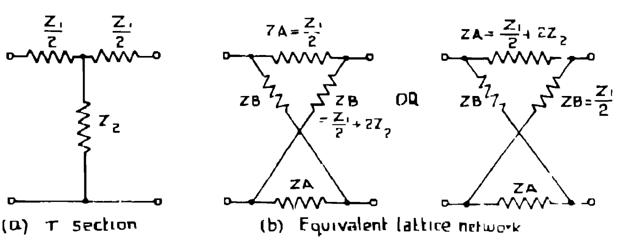


FIG. 704.—Equivalence of T section and lattice section.

In Fig. 704 is shown the equivalent lattice section for a normal T section of the ladder type. Clearly this equivalence is not complete, for the lattice section is inherently a balanced section, whereas the T section is not. They are, however, equivalent from the point of view that they have the same characteristic impedance and the same propagation constant. A mathematical proof of this will be found in Chapter 13, but the equivalence may be verified quite simply as follows:

Since $Z_0 = \sqrt{Z_{SC} \cdot Z_{OC}}$ and $\tanh \gamma = \sqrt{\dfrac{Z_{SC}}{Z_{OC}}}$ for any network,

it follows that if two networks have the same open-circuit and

short-circuit impedances, they must therefore have the same values of characteristic impedance and propagation constant.

Considering the open-circuit impedance :—

For the T section.— $$Z_{OC} = \frac{Z_1}{2} + Z_2 \tag{104}$$

For the lattice section $$Z_{OC} = \frac{1}{2}(Z_A + Z_B) \tag{105}$$

If these are the same.—

$$Z_A + Z_B = Z_1 + 2Z_2 \tag{106}$$

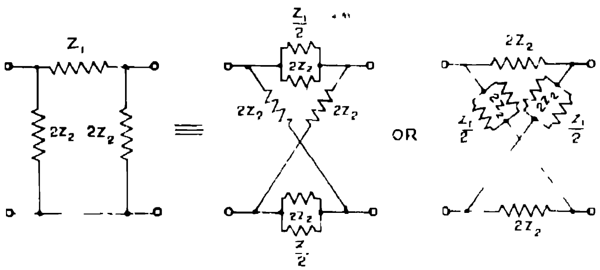


FIG. 705.—Equivalence of π section and lattice section.

Considering the short-circuit impedance :—

For the T section.— $$Z_{SC} = \frac{Z_1}{2} + \frac{Z_1 Z_2}{2\left(\frac{Z_1}{2} + Z_2\right)}$$

i.e. $$Z_{SC} = \frac{\frac{Z_1^2}{4} + Z_1 Z_2}{\frac{Z_1}{2} + Z_2} \tag{107}$$

and for the lattice section. $$Z_{SC} = \frac{2Z_A Z_B}{Z_A + Z_B} \tag{108}$$

If these are the same :—

$$\frac{2Z_A Z_B}{Z_A + Z_B} = \frac{\frac{Z_1^2}{4} + Z_1 Z_2}{\frac{Z_1}{2} + Z_2}$$

or using equation (106) :—

$$Z_A Z_B = \frac{Z_1^2}{4} + Z_1 Z_2 \tag{109}$$

Clearly the solution of equations 106 and 109 is :—

$$Z_A = \frac{Z_1}{2} \quad \text{or} \quad \frac{Z_1}{2} + 2Z_2 \tag{110}$$

and

$$Z_B = \frac{Z_1}{2} + 2Z_2 \quad \text{or} \quad \frac{Z_1}{2} \tag{111}$$

In the same way the equivalence between a π section and a lattice section may be established, as in Fig. 705.

This equivalence may be used to derive the lattice equivalent of the ladder prototype low-pass filter ; other ladder filters may be treated in the same way if their lattice equivalents are required.

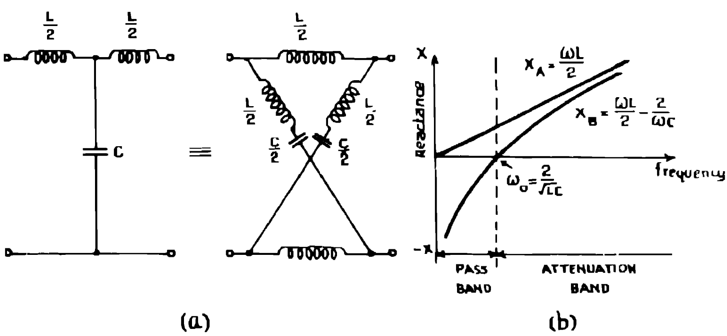


FIG. 706.—Lattice equivalent of prototype T section low-pass filter.

Fig. 706*a* shows a prototype T low-pass filter section having a cut-off frequency $\frac{1}{\pi\sqrt{LC}}$, and its lattice equivalent. Fig. 707*a* shows the corresponding π low-pass filter section, and its lattice equivalent.

If a lattice filter section of unknown type is encountered it may

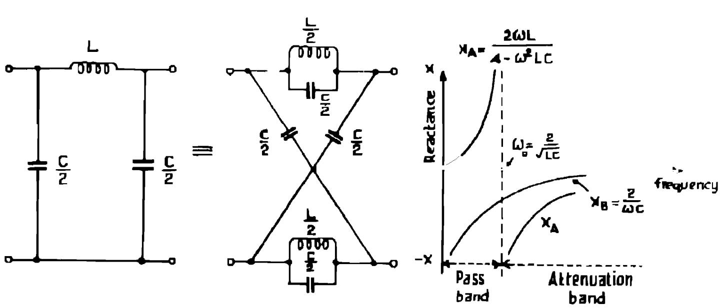


FIG. 707.—Lattice equivalent of prototype π section low-pass filter.

be explored by the reactance sketch method. For in the general lattice network of Fig. 704 :—

$$Z_0 = \sqrt{Z_A Z_B} \tag{112}$$

Thus if Z_A and Z_B are reactances of opposite sign, Z_0 is a pure resistance and the filter has zero attenuation. If, on the other hand, Z_A and Z_B are reactances of the same sign, then Z_0 is a pure reactance and the filter attenuates.

Considering the lattice filter sections of Figs. 706*a* and 707*a*, the corresponding reactance sketches are shown in Figs. 706*b* and 707*b*. In both cases it will be seen from the reactance sketches that the filters are low-pass, having a cut-off frequency $f_0 = \dfrac{1}{\pi\sqrt{LC}}$.

Insertion loss of a lattice section filter

Like any other 4-terminal network, the insertion loss and insertion phase-shift of a lattice filter may be determined, using equations 7 and 8 of Chapter 13 (page 566).

When, however, such a network is inserted between purely resistive impedances (such as, say, the filter's design impedances), the problem is best approached from first principles.

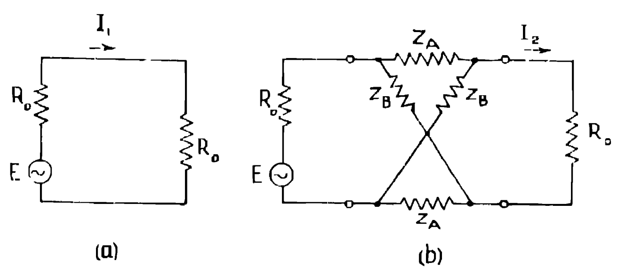


FIG 708.– Calculation of the insertion loss of a lattice section

Consider a generator of EMF E and internal impedance R_0 working into load R_0 (Fig. 708*a*). Let the current flowing be I_1.

If a lattice network now be inserted between the generator and the load (Fig. 708*b*), let the new current flowing be I_2.

Applying Kirchhoff's laws to Fig. 708*b*, it may be shown that :—

$$I_2 = -\frac{E}{2} \cdot \frac{Z_A - Z_B}{(R_0 + Z_A)(R_0 + Z_B)}$$

Considering Fig. 708*a* :—

$$I_1 = \frac{E}{2R_0}$$

Hence $$\frac{I_1}{I_2} = -\frac{(R_0 + Z_A)(R_0 + Z_B)}{R_0 (Z_A - Z_B)}$$

Let $\quad Z_A = jX_A \quad \text{and} \quad Z_B = jX_B$

Then $$\frac{I_1}{I_2} = -\frac{(R_0 + jX_A)(R_0 + jX_B)}{jR_0(X_A - X_B)}$$

$$\frac{I_1}{I_2} = -\frac{R_0(X_A + X_B) - j(R_0^2 - X_AX_B)}{R_0(X_A - X_B)} \tag{113}$$

Consider first the modulus of this expression, which determines the insertion loss

$$\left|\frac{I_1}{I_2}\right| = \frac{\sqrt{R_0^2(X_A + X_B)^2 + (R_0^2 - X_AX_B)^2}}{R_0(X_A - X_B)}$$

$$= \frac{\sqrt{(R_0^2 + X_A^2)(R_0^2 + X_B^2)}}{R_0(X_A - X_B)}$$

$$= \frac{1}{\dfrac{X_A}{\sqrt{R_0^2 + X_A^2}} \cdot \dfrac{R_0}{\sqrt{R_0^2 + X_B^2}} - \dfrac{R_0}{\sqrt{R_0^2 + X_A^2}} \cdot \dfrac{X_B}{\sqrt{R_0^2 + X_B^2}}}$$

$$= \frac{1}{\sin\theta_1 \cos\theta_2 - \cos\theta_1 \sin\theta_2}$$

$$= \frac{1}{\sin(\theta_1 - \theta_2)}$$

where $\theta_1 = \tan^{-1}\dfrac{X_A}{R_0}$ and $\theta_2 = \tan^{-1}\dfrac{X_B}{R_0}$

(a) (b)

FIG 709

The insertion loss is given by :—

$$20\log_{10}\left|\frac{I_1}{I_2}\right| = 20\log_{10}\frac{1}{\sin(\theta_1 - \theta_2)} \text{ db} \tag{114}$$

or $$\log_e\left|\frac{I_1}{I_2}\right| = \log_e\frac{1}{\sin(\theta_1 - \theta_2)} \text{ nepers} \tag{115}$$

Considering the phase angle of expression 113, it is seen that :—

$$\beta = \tan^{-1} - \frac{R_0^2 - X_AX_B}{R_0(X_A + X_B)}$$

$$= \tan^{-1} - \frac{1 - \dfrac{X_A}{R_0}\dfrac{X_B}{R_0}}{\dfrac{X_A}{R_0} + \dfrac{X_B}{R_0}}$$

$$= \tan^{-1} - \frac{1 - \tan\theta_1 \tan\theta_2}{\tan\theta_1 + \tan\theta_2}$$

$$\therefore \qquad \beta = \tan^{-1} \frac{-1}{\tan(\theta_1 + \theta_2)}$$

Hence the insertion phase-shift β is given by :—

$$\beta = \theta_1 + \theta_2 \pm (2n + 1)\frac{\pi}{2} \tag{116}$$

Both equations 114 and 115, and equation 116, are great simplifications on equations 7 and 8 of Chapter 13 (page 566), which are the equations employed for the determination of insertion losses and insertion phase angles of ladder filters. It will be noted, therefore, that in many cases it is convenient to convert a ladder filter into its corresponding lattice, so that the above insertion loss equations may be employed.

Equivalent bridged-T sections

As may be seen by studying the low-pass filter sections of Figs. 706*a* and 707*a*, the main disadvantage of lattice sections is in

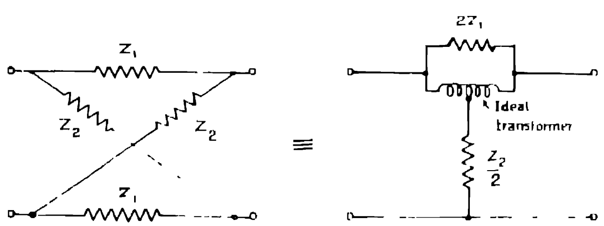


FIG. 710.—Bridged T equivalent of a lattice section (using transformer).

the greater number of components employed ; this may be obviated to a certain extent by converting from a lattice section to a bridged-T section as shown below.

From Figs. 711*a* and *b*, it will be seen that, for the bridged-T section :—

$$Z_{OO} = \frac{Z_2}{2} + \frac{1}{4} \cdot 2Z_1 = \frac{Z_1 + Z_2}{2} \tag{117}$$

and

$$Z_{SO} = \frac{2Z_1 \times 4 \cdot \frac{Z_2}{2}}{2Z_1 + 4 \cdot \frac{Z_2}{2}} = \frac{2Z_1 Z_2}{Z_1 + Z_2} \tag{118}$$

By comparison with equations 105 and 108 the equivalence of Fig. 710 is established. Using this equivalence, lattice low-pass sections may be replaced by bridged-T low-pass sections as in Fig. 712.

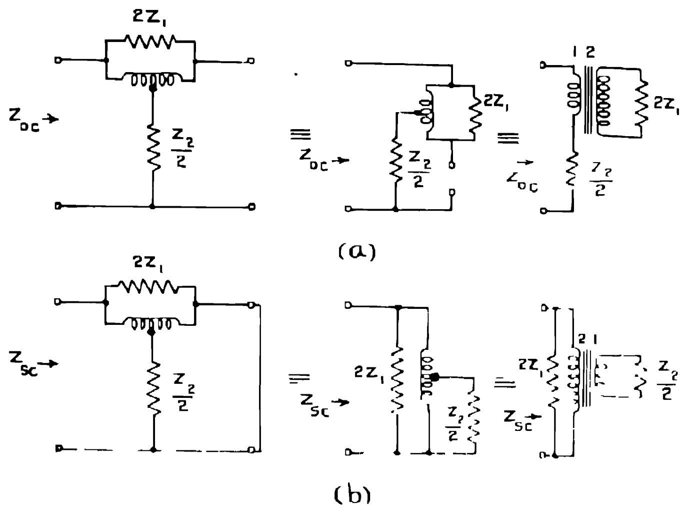


FIG. 711.—Equivalent circuit of bridged-T on open and short circuit.

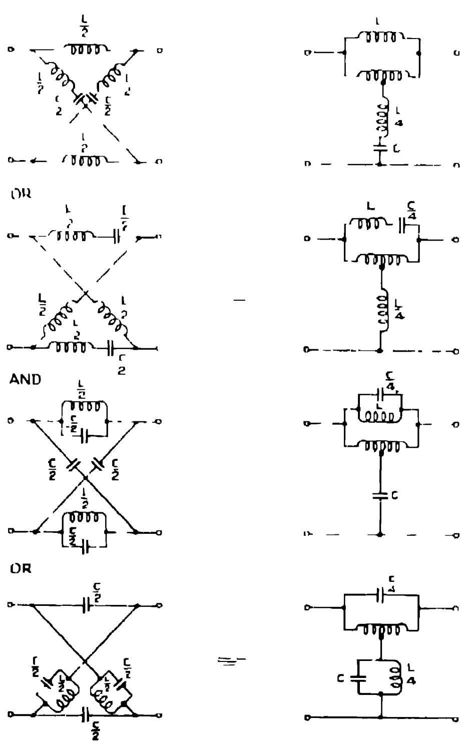


FIG. 712.—Bridged-T equivalents of lattice low-pass filter sections.

The limitations on such filter sections are imposed by the discrepancies introduced by substituting a practical transformer for the ideal transformer.

CRYSTAL FILTERS

Certain substances, notably quartz, exhibit what is known as the "piezo-electric" effect; that is to say, a mechanical strain applied to a suitably cut piece of the substance causes an EMF to be developed between two surfaces of that piece; and conversely, an EMF applied between two faces of the piece cause a mechanical deformation. These effects are used in piezo-electric microphones and similar devices.

Slices cut from a quartz crystal also have another important property: they behave electrically as a resonant circuit, of extremely high Q, resonant at the natural frequency of mechanical vibration of the slice.

The complete quartz crystal in natural state is of the general shape shown in Fig. 713*a*, and for convenience of reference, three

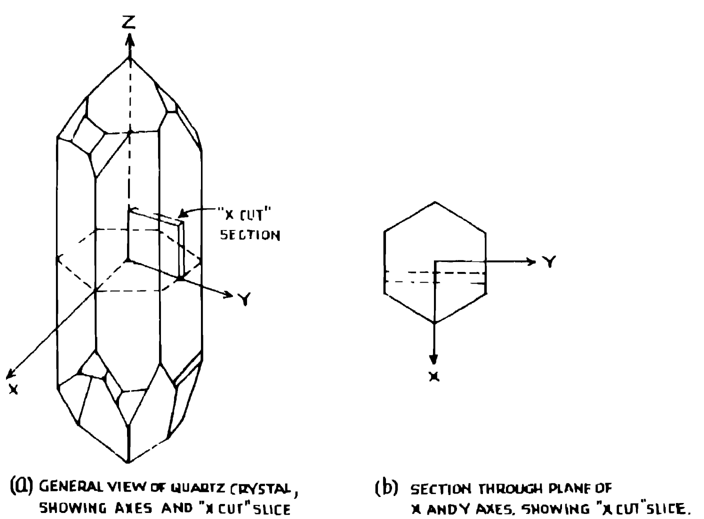


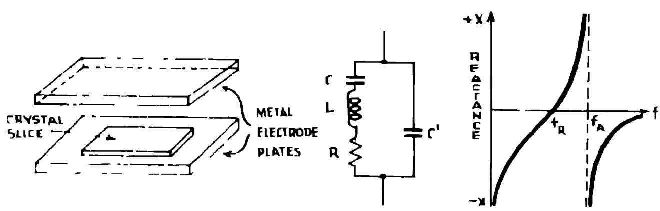
FIG. 713.—Quartz crystal showing "X cut" slice.

axes through the crystal are defined as follows: the vertical axis passing through both "points" of the crystal is called the optical or Z-axis; the line at right angles to the Z-axis parallel to any major face of the crystal is called the electrical or X-axis; and the line at right angles to both these axes, and perpendicular to any face of the crystal, is called the mechanical or Y-axis. Slices cut from the crystal for various purposes may then be described in terms of the angles between them and the three axes of the crystal; thus a slice cut with its faces perpendicular to the X-axis is called

an "X cut" crystal, as shown in Fig. 713, and this "cut" frequently used in filters.

If such a slice be mounted between two flat metal plates which electrical connection can be made (*see* Fig. 714*a*), it is found to behave like the equivalent circuit given in Fig. 714*b*. C, L, and R are merely the electrical equivalents of corresponding mechanical properties of the crystal, and their values are determined by the dimensions and physical constants of the crystal material. C' is the electrical capacity between the faces of the crystal, and in practice its effective value may be increased by the capacity of the wiring.

The series circuit composed of C, L, and R resonates at the natural (mechanical) resonant frequency of the crystal; let this frequency be f_R. In addition to this series resonant frequency, the circuit of Fig. 714*b* also has an anti-resonant frequency; let this be f_A. The reactance sketch for the crystal is therefore of the



(a) (b) (c)

FIG. 714.—Equivalent circuit of quartz crystal.

(*a*) Crystal in "solid" type of mounting. (*b*) Equivalent circuit of quartz crystal. (*c*) Reactance sketch for quartz crystal.

form shown in Fig. 714*c*. Neglecting the equivalent resistance R (which, owing to the high Q, is relatively small), the impedance of the circuit of Fig. 714*b* is :—

$$Z = \frac{\dfrac{-j}{\omega C'}\left(j\omega L - \dfrac{j}{\omega C}\right)}{j\omega L - \dfrac{j}{\omega C} - \dfrac{j}{\omega C'}}$$

$$= \frac{-j\,(\omega^2 LC - 1)}{\omega CC'\left(\omega^2 L - \dfrac{C + C'}{CC'}\right)}$$

Thus the series resonant frequency is given by :—

$$\omega_R{}^2 = \frac{1}{LC}, \quad \text{or} \quad f_R = \frac{1}{2\pi\sqrt{LC}} \qquad (119,$$

and the anti-resonant frequency by :—

$$\omega_A{}^2 = \frac{1}{L \cdot \dfrac{CC'}{C + C'}}$$

$$\therefore \qquad f_A = \frac{1}{2\pi\sqrt{L \cdot \dfrac{CC'}{C + C'}}}$$

$$= \frac{\sqrt{1 + \dfrac{C}{C'}}}{2\pi\sqrt{LC}}$$

$$= f_R \cdot \sqrt{1 + \frac{C}{C'}} \qquad (120)$$

This can be expanded by the Binomial Theorem as :—

$$f_A = f_R\left(1 + \frac{C}{C'}\right)^{\frac{1}{2}}$$

$$= f_R\left\{1 + \tfrac{1}{2} \cdot \frac{C}{C'} + \frac{(\frac{1}{2}) \cdot (-\frac{1}{2})}{1 \cdot 2} \cdot \left(\frac{C}{C'}\right)^2 + \ldots\right\}$$

For the crystal itself, with no stray capacities, C' is normally 125 times C, so that the ratio $\frac{C}{C'}$ is $\frac{1}{125}$. Terms containing second and higher powers of $\frac{C}{C'}$, can therefore be neglected, so that :

$$f_A = f_R\left(1 + \frac{C}{2C'}\right) \simeq 1 \cdot 004 \, f_R \qquad (121)$$

The separation between the resonant and anti-resonant frequencies is therefore :

$$f_A - f_R = f_R\left(1 + \frac{C}{2C'}\right) - f_R$$

$$\text{i.e.} \qquad f_A - f_R = f_R \cdot \frac{C}{2C'} = 0 \cdot 4 \text{ per cent. of } f_R \qquad (122)$$

Stray capacity in parallel with the crystal has the effect of increasing the effective value of C', so that $\frac{C}{C'} < \frac{1}{125}$; it therefore lowers the anti-resonant frequency f_A and reduces the separation $(f_A - f_R)$.

If the dimensions of an X-cut quartz crystal slice be denoted by x, y, and z (in centimetres) in the directions of the X, Y, and Z axes respectively, the component values of the equivalent circuit

can be found approximately from the following empirical formulæ

$$L = 115\frac{xy}{z} \quad \text{Henries} \qquad (123)$$

$$C = 0\cdot0032\frac{yz}{x} \quad \mu\mu\text{F} \qquad (124)$$

$$C' = 0\cdot40\frac{yz}{x} \quad \mu\mu\text{F} \qquad (125)$$

No useful formula can be given for R; for although the actual value of R for the crystal may be very low, its effective value is considerably increased in practice by the mechanical damping of the crystal mounting. The inherent Q of a crystal slice is almost infinite; but even with the most careful mounting, in vacuo, with sprayed gold electrodes, this is reduced to several hundred thousand and with normal mountings, it is of the order of 5000 to 20,000. Even these values, however, are appreciably higher than those obtainable from circuits using normal type inductances and capacities, which seldom yield a Q higher than about 200.

Example.—

A quartz crystal slice measures 0·2, 2·5, and 0·5 cm in the directions of the X, Y, and Z axes respectively (*see* Fig. 715*a*).

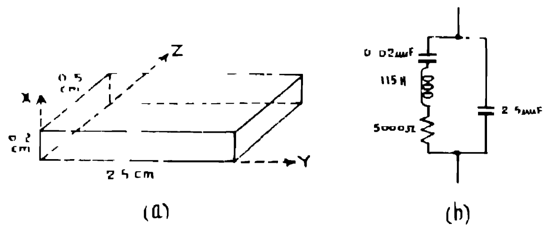


FIG. 715.—Dimensions and equivalent circuit of an actual crystal.

Find its equivalent circuit and its resonant and anti-resonant frequencies.

From equations 123, 124, and 125:—

$$L = 115\frac{xy}{z} = 115 \,.\, \frac{0\cdot2 \,.\, 2\cdot5}{0\cdot5} = 115\text{ H}$$

$$C = 0\cdot0032\frac{yz}{x} = 0\cdot0032 \,.\, \frac{2\cdot5 \,.\, 0\cdot5}{0\cdot2} = 0\cdot02\ \mu\mu\text{F}$$

$$C' = 0\cdot40\frac{yz}{x} = 0\cdot40 \,.\, \frac{2\cdot5 \,.\, 0\cdot5}{0\cdot2} = 2\cdot5\ \mu\mu\text{F}$$

$$f_R = \frac{1}{2\pi\sqrt{LC}} = \frac{1}{2\pi\sqrt{115 \,.\, 0\cdot02 \,.\, 10^{-12}}} = 104\cdot8\text{ kc/s}$$

$$f_A = \frac{1}{2\pi\sqrt{L \cdot \frac{CC'}{C+C'}}} = \frac{1}{2\pi\sqrt{115 \cdot \frac{0\cdot 02 \cdot 2\cdot 5 \cdot 10^{-12}}{2\cdot 52}}} = 105\cdot 2 \text{ kc/s}$$

This particular crystal slice was found to have a Q of 15,000, corresponding to an equivalent resistance of: –

$$R = \frac{\omega L}{Q} = \frac{2\pi \cdot 104\cdot 8 \cdot 10^3 \cdot 115}{15,000} = 5000\,\Omega.$$

The equivalent circuit is then as shown in Fig. 715*b*.

Single-crystal filter

The simplest method of employing a crystal in a filter circuit is shown in Fig. 716*a*. Here the behaviour of the crystal may be determined by considering it to be replaced by its equivalent circuit (*see* Fig. 714*b*); thus it presents a low impedance at its resonant frequency, an almost infinite impedance at its anti-resonant frequency, and a high impedance at all other frequencies. The attenuation-frequency curve for this filter is therefore of the form shown in Fig. 716*b*.

The anti-resonant frequency f_A can be altered by varying the effective shunt capacity across the crystal (C' in Fig. 714*b*). This can conveniently be done by using a centre-tapped input transformer and a variable "balancing" condenser C_B, as shown in

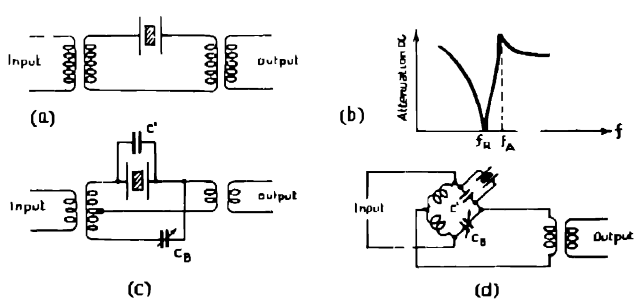


FIG 716 — Single-crystal band-pass filter.

Fig. 716*c*, which can be looked upon as a bridge circuit as in Fig. 716*d*. If C_B be adjusted to zero capacity, the circuit is essentially the same as Fig. 716*a*, but with only half the input voltage utilised.

If C_B then be increased, some current will reach the output through it, approximately in anti-phase to that passing *via* C'; the resulting current reaching the output is therefore less than that passing *via* C', and the effect is the same as would be obtained if C' had been reduced. Anti-resonance occurs at that frequency for which the capacitive current passing *via* C' is equal in magnitude to the

inductive current passing *via* C, L, and R; reduction of the current reaching the output *via* C' therefore causes the anti-resonance to occur at a frequency for which the inductive current through C, L, and R is less—*i.e.* at a higher frequency.

If C_B is increased until it is equal to C', no anti-resonance will occur; while if it be further increased, the current reaching the output *via* C_B will be greater than that *via* C', and the anti-resonance will occur at a frequency for which the current reaching the output *via* C, L, and R is capacitive—*i.e.*, at a frequency *below* f_R. Thus as C_B is increased from zero, the anti-resonant frequency f_A is first increased from its original value, and reaches infinity when $C_B = C'$; further increase of C_B raises f_A from zero frequency until it approaches f_R when C_B is made very large.

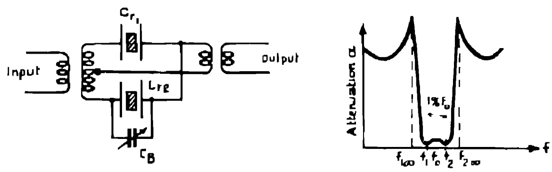
A filter of this type is especially useful when it is desired to transmit one particular frequency, and to attenuate a neighbouring frequency severely. For most communications purposes, however, comparatively wide bands of frequencies have to be passed, and for this purpose the response curve of Fig. 716*b* is unsuitable; for the frequency band over which the attenuation is low is extremely narrow. The width of the response curve can be modified to a certain extent by varying the impedance of the input and output circuits; for maximum sharpness (*i.e.*, highest Q and minimum bandwidth) both input and output impedances should be low but for maximum output voltage at the resonant frequency the output impedance should be high. Even when its sharpness is reduced by increasing the input and output impedances, the response of a filter of this type is unsatisfactory for many purposes owing to its pointed nature. Various types of filter have therefore been developed in which advantage is taken of the extremely high Q of quartz crystals, to furnish a band-pass filter with a very low and constant attenuation in the pass bands, but with a sharpness of cut-off approaching that of "ideal" resistanceless filters.

Double-crystal band-pass filter

A filter circuit giving a pass band up to about 1 per cent. of the mid-band frequency is shown in Fig. 717*a*. The resonant frequencies of the two crystals Cr_1 and Cr_2 are at the lower and upper edges of the pass band respectively, while the input and output circuits are tuned to the mid-band frequency. At frequencies below the lower edge of the pass band, both crystals present a capacitive impedance; owing to the "push-pull" input and parallel output connections, the outputs from the two crystals are in anti-phase, so that the resultant output is less than that from one crystal alone. Between the resonant frequencies of the two crystals, Cr_1 presents an inductive and Cr_2 a capacitive impedance; but owing to the method of connection, the two outputs are additive, so that their combined output is greater than that obtained from either crystal alone. Above the higher edge of the pass band, both crystals present an inductive impedance, but

their outputs, being in anti-phase, tend to cancel out. Thus a low attenuation is obtained between the resonant frequencies (f_1 and f_2) of the two crystals, and a high attenuation outside this band.

If a balancing condenser C_B is so adjusted that the shunt capacities across the two crystals are equal, no anti-resonance occurs. But if this condenser be so adjusted that the shunt capacity across the higher-frequency crystal Cr_2 exceeds that across Cr_1 an anti-resonance occurs just above the resonant frequency f_2 of Cr_2; and a second anti-resonance occurs just below the resonant frequency f_1 of the lower-frequency crystal Cr_1, as explained above in connection with the single-crystal filter. The attenuation-frequency curve then obtained is as shown in Fig. 717*b*, and it will



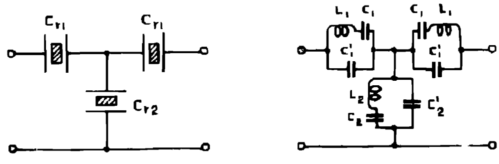
(a) (b)

FIG 717.—Double crystal band pass filter

be noted that there is an increase in attenuation at the centre of the pass band, the wider the pass band, the more serious is this increase. For this reason, filters of this type are used only when the bandwidth required is less than about 1 per cent. of the mid-band frequency; thus for a bandwidth of 3000 c/s, this type of filter would be satisfactory for mid-band frequencies above 300 kc/s. In line communication, however, crystals are employed chiefly as the reactive elements in band-pass filter circuits of conventional type.

T type crystal filters

The T type crystal filter forms a good example of the use of crystals as the reactive elements in conventional filter circuits,



(a) (b)

FIG. 718.—T type crystal band-pass filter.

and its analysis shows clearly how the reactance sketches for crystals used in this way are manipulated. This type of filter is shown in Fig. 718*a*, and its equivalent circuit (neglecting resistance) in Fig. 718*b*.

The crystals Cr_1 in the series arms are carefully selected to have a series resonant frequency exactly equal to the anti-resonant frequency of the crystal in the shunt arm. Let the resonant and anti-resonant frequencies of the series arms be f_{R1} and f_{A1} respectively, and those of the shunt arm f_{R2} and f_{A2}; then $f_{R1} = f_{A2}$. Let the reactance of the series arm be X_1, and that of the shunt arm be X_2; the variation with frequency of X_1 and X_2 is then as shown in Fig. 719*a*. From this can be drawn the reactance sketch for the filter, as in Fig. 719*b*.

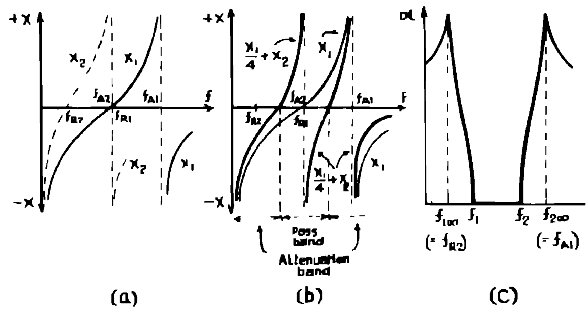
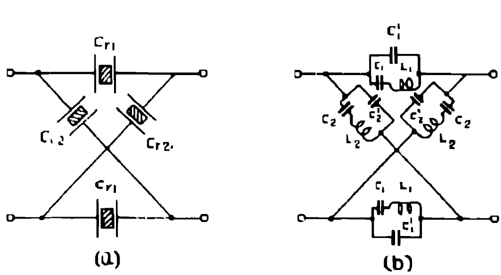


FIG. 719.—Reactance sketch for a T type crystal band-pass filter.

It will be seen that the filter has a band-pass characteristic, with a pass band from f_1 to f_2, and frequencies of infinite attenuation at $f_{1\infty} = f_{R2}$ and $f_{2\infty} = f_{A1}$. The attenuation-frequency characteristic, shown in Fig. 719*c*, is seen to be of the same form as that of an *m*-derived band-pass section; while the bandwidth $(f_2 - f_1)$ is seen to be less than $(f_{A1} - f_{R2})$ - *i.e.*, less than 0·8 per cent. of the mid-band frequency.

Lattice type crystal filters

Crystal filters are frequently made up in lattice form, as shown in Fig. 720*a*, with the equivalent circuit as in Fig. 720*b*.



Lattice crystal filter, with equivalent circuit.

FIG. 720.—Lattice crystal band-pass filter.

The crystals Cr_1 are a carefully matched pair, as also are Cr_2. The crystals are so cut that the series resonant frequency of one pair corresponds with the anti-resonant frequency of the other. Using the method set out on page 686, the reactance sketch is shown in Fig. 721*a*, indicating that the filter has a band-pass characteristic. The pass band is seen to extend from $f_1 = f_{R1}$ to $f_2 = f_{A2}$, so that the bandwidth is 0·8 per cent of the mid band frequency $f_0 = f_{A1} = f_{R2}$. The corresponding attenuation-frequency characteristic is shown in Fig. 721*b*.

By suitable choice of the crystals in the series and lattice arms, the reactance curves may be made to cross as shown in Fig. 721*c*. The points of intersection correspond to equal impedances in

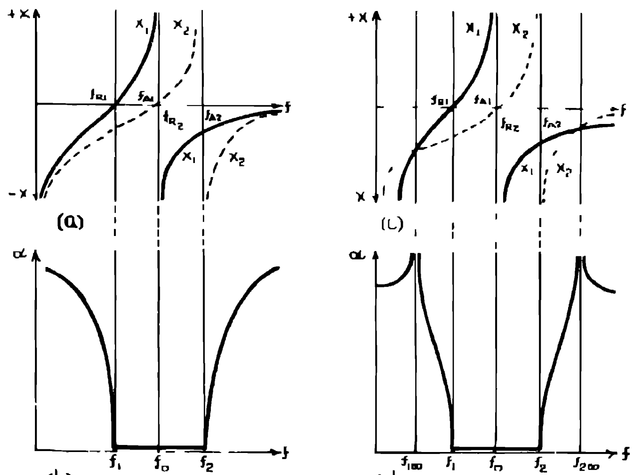


FIG. 721.—Reactance sketch for a lattice crystal band-pass filter.

the series and lattice arms giving rise to frequencies of infinite attenuation $f_{1\infty}$ and $f_{2\infty}$, as shown in Fig. 721*d*, since the load would now be across the diagonal of a balanced bridge. Such an arrangement gives considerable discrimination against frequencies immediately outside the pass band.

In lattice crystal filter circuits a pair of identical crystals are required in the two series arms, and a second pair in the two lattice arms. In practice, each such pair of identical crystals can be replaced by a single crystal slice with two independent pairs of terminals. Each face of the slice carries two metal coatings, usually of aluminium, insulated from each other, the two coatings on one end of the crystal slice being used for one arm of the circuit, and those on the other end for the opposite arm. The whole slice vibrates mechanically, but owing to the symmetrical disposition of the two identical crystals in a lattice filter section, no undesired

coupling between different parts of the circuit can occur. The use of one crystal slice to perform the two electrical functions in this way is not only economical and convenient, but also ensures that the two " opposite " crystals in the filter do, in fact, have identical characteristics.

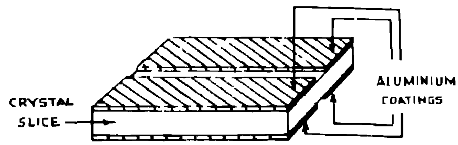


FIG. 722.—Use of a single crystal slice to provide a matched pair of crystal elements.

Crystals with series inductance

It has been seen that the bandwidth obtainable from a simple ladder or lattice filter using crystals as the reactive elements is limited, by the proximity of the resonant and anti-resonant frequencies, to about 1 per cent. of the mid-band frequency. It has also been seen that these frequencies can be brought nearer together by the addition of shunt capacity across the crystal, and that they can be separated by reducing the effective shunt capacity. There is, however, an irreducible minimum value to the shunt capacity of the crystal (C' in Fig. 714b) ; and in the case of a ladder or lattice type filter section, this capacity cannot conveniently be " balanced out ", as it can in the case of the double-crystal filter shown in Fig. 717. The resonant and anti-resonant frequencies of a crystal can, however, be separated by the addition of inductance in series or in shunt with the crystal.

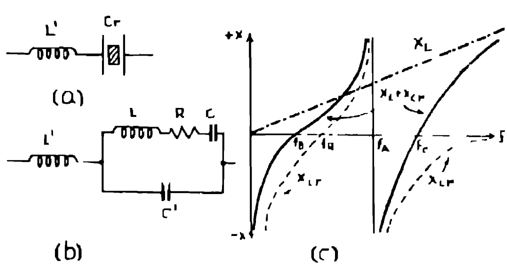


FIG. 723.—Effect of adding inductance in series with a crystal.

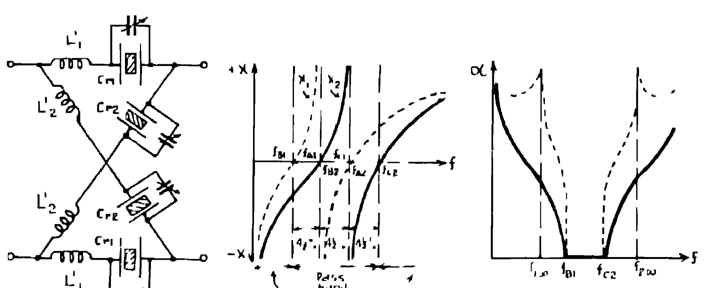
If an inductance L' be connected in series with a crystal Cr, as shown in Fig. 723a, the equivalent circuit is as given in Fig. 723b. The reactance X_{Cr} of the crystal is given by the broken lines in Fig. 723c, with resonant and anti-resonant frequencies f_B and f_A ; while the reactance X_L of the added inductance is given by the chain-dotted line. The sum of these two gives the total reactance,

shown by the full-line curve, and is seen to exhibit an anti-resonance at f_A, and *two* series resonances, at f_B and f_C. The separation between f_B and f_A is clearly greater than that between f_B and f_A; it can be shown that maximum bandwidth is obtained if the value of L' is so chosen that $(f_A - f_B) = (f_C - f_A)$, and in this case the separation $(f_A - f_B)$ is $4\frac{1}{2}$ per cent. of f_A. Thus from the point of view of maximum bandwidth, there is an optimum value of L' for any crystal.

Since any coil has an appreciable resistance, and a low value of Q compared with that of the crystal, it is at once apparent that the addition of an inductance L' in series with a crystal will adversely affect the performance of a filter. In certain cases, however, the added resistance can be made to appear outside the arms of the filter, and be combined with an external resistance to form an attenuator, which will give a constant increase in attenuation at all frequencies (*i.e.*, in both the pass and the attenuation bands), the increase being usually only a few decibels. This can best be seen by considering the addition of series inductances to a lattice filter section.

Broad-band lattice filter

Fig. 724*a* shows a lattice filter section with inductances L_1', L_2', added in series with each of the four crystals. The anti-resonant frequency f_{A1} of the series arms is made equal to the lower



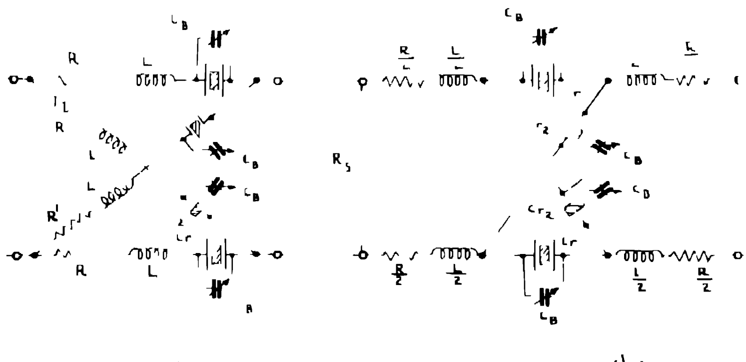
(a) (b) (c)

FIG. 724.—Broad-band lattice filter using series inductances.

series resonant frequency f_{B2} of the lattice arms; and the anti-resonant frequency f_{A2} of the lattice arms is made equal to the higher series resonant frequency f_{C1} of the series arms. The reactance sketch for this section is therefore as shown in Fig. 724*b*, and the attenuation-frequency characteristic as in Fig. 724*c*. The component values can, of course, be so chosen that the reactance curves X_1 and X_2 cross, giving rise to frequencies of infinite attenuation $f_{1\infty}$ and $f_{2\infty}$, and a response of the form shown dotted in Fig. 724*c*.

Since each of the three separations $(f_{A1} - f_{B1})$, $(f_{01} - f_{A1})$, $(f_{A2} - f_{B2})$, and $(f_{02} - f_{A2})$ is approximately $4\frac{1}{2}$ per cent of the mid-band frequency, the total bandwidth is $3 \times 4\frac{1}{2} = 13\frac{1}{2}$ per cent of the mid-band frequency. If this is too wide it may be reduced by connecting condensers in parallel with the crystals, and by this means any bandwidth from $13\frac{1}{2}$ per cent down to $\frac{1}{2}$ per cent of the mid-band frequency may be obtained. For bandwidths less than $\frac{1}{2}$ per cent, the loss caused by the resistance of the inductances becomes excessive, but below 1 per cent bandwidth the all-crystal filter can be used. For bandwidths greater than 13 per cent ordinary inductance and condenser filters are satisfactory.

The inductances L_1' and L_2' in the series and lattice arms may all be made equal and the anti-resonant frequencies adjusted to equality with the appropriate resonant frequencies by means of condensers C_B in parallel with the crystals as shown in Fig. 725a, where the inherent series resistance of each coil is shown as R'. Then by the theorem given in Chapter 13 (page 598) the impedance formed by L' and R' in series can be subtracted from all four arms of the lattice and placed in series with the input or output terminals



(a) (b)

Fig. 725. Showing how resistance of series inductances in lattice filter can be made to form an attenuator independent of frequency.

of the section as shown in Fig. 725b. If a resistance R_s be connected across the terminals of the resulting network it will form, together with the resistances R', an attenuator, and the value of R_s can be chosen to give correct impedance matching and minimum loss.

In a similar manner a wide-band filter can be obtained by connecting inductances in *parallel* with each crystal, and these added inductances, if equal, can be removed from the arms of the lattice and placed in parallel with the input or output terminals of the resulting network. A broad band crystal filter with series inductances has a low characteristic impedance (of the order of 100Ω), while the corresponding filter with inductances added in parallel with the crystals has a high characteristic impedance, of the order of 100,000Ω.

CHAPTER 16

LINE TRANSMISSION

Line transmission is the theory of the propagation of electric waves along transmission lines. These transmission lines are assumed to consist of a pair of wires that are uniform throughout their whole length. Provided that this uniformity holds good, it is immaterial for the general theory whether the two wires are air-spaced on telegraph poles, are two conductors in an underground cable, or form a pair in a field quad cable.

Only steady-state currents and voltages will be considered, and the problem resolves itself into one of finding the current and voltage at any point along the length of the line, when a known voltage (or current) is continuously applied to the sending end.

In this chapter, the problem will be treated as far as possible in a non-mathematical manner, a more rigid mathematical treatment will be given in Chapter 17.

THE INFINITE LINE

The propagation of electric waves along any uniform and symmetrical transmission line may be deduced in terms of the results for a hypothetical line of infinite length having electrical constants per unit length identical to those of the line under consideration. For this reason, the propagation of electric waves along an infinite line will be considered first.

FIG. 726.—Infinite line.

When an alternating voltage is applied to the sending end of an infinite line (*see* Fig. 726), a finite current will flow due to the capacity and the leakage conductance between the two wires constituting the line. The value of this current will depend upon these two factors and others to be investigated later.

The ratio of the voltage applied, to the current flowing, will give the input impedance. This input impedance is known as the "characteristic impedance" of the line, and is denoted by Z_0.

The characteristic impedance of any line is defined as the impedance looking into an infinite length of the line (*c.f.* characteristic impedance of a network, Chapter 13, page 561).

This characteristic impedance, in the case of a telephone line, is a vector quantity having a modulus $|Z_0|$ (usually between 200 and 1600 ohms), and an angle φ (usually lying between $-45°$ and $0°$). Since this angle lies in the fourth quadrant, it may be written either as $\angle -\varphi°$ or as $\overline{\diagdown \varphi°}$. The modulus and angle of the characteristic impedance will, in general, vary with frequency; and the frequency at which the impedance has been measured should be stated. The following table shows the characteristic impedance of some typical army cables measured at 1600 c/s :—

TABLE XIX

Type of Cable	Z_0
Cable, electric, D 8, twisted	$500 \angle -32°$
Cable, electric, D 8, single, 9 in spacing	$1600 \angle -30°$
Field Quad (unloaded)	$300 \angle -37°$

In addition to possessing an input impedance Z_0, an infinite line has the following important properties :-

(1) Since the line has infinite length, no waves will ever reach the distant end, hence there will be no possibility of reflection at the distant end and no reflected waves will return to the sending end.

(2) For the same reason, when a voltage is applied to the sending end, the current flowing will depend only on the characteristic impedance Z_0, and will be unaffected by the terminating impedance Z_R at the distant end.

It may be noted that in practice this last condition is approximately fulfilled by many long lines.

Short line terminated in Z_0

Consider an infinite line having input terminals 1 and 2 (*see* Fig. 727*a*). The impedance looking in at terminals 1 and 2 will, by definition, be Z_0. Suppose that a short section AB at the near end of the line is now removed (Fig. 727*b*), so that the line now starts at terminals 3 and 4. The impedance looking in at terminals 3 and 4 will still be Z_0, since the removal of the short section does not affect the infinite nature of the line. This means that the short section AB, from an electrical point of view, was originally terminated in an impedance Z_0 at B. If the short section AB is now terminated in an actual impedance Z_0, the current and voltage at all points along its length will be exactly the same as if it were terminated in an infinite length of line.

It therefore follows that a short line terminated in Z_0 behaves

electrically at all points along its length, as if it were an infinite line. In particular, this means that the input impedance will be Z_0, and that there will be no reflection. Before a line can be thus

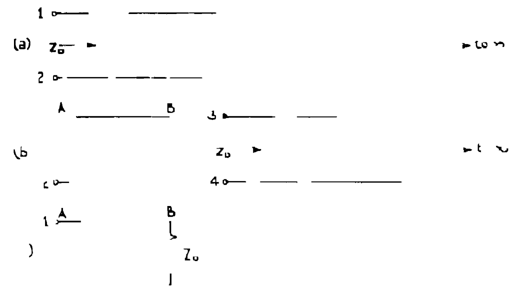


FIG. 727. Short line terminated in Z_0

terminated however a method must be developed for the practical determination of Z_0 when only a short length of line is available.

Determination of Z_0 for a short line

A short line may be considered as a complex electrical network, and like any other network it may at the frequency under consideration be represented by a T section (see Fig. 728a). If the short

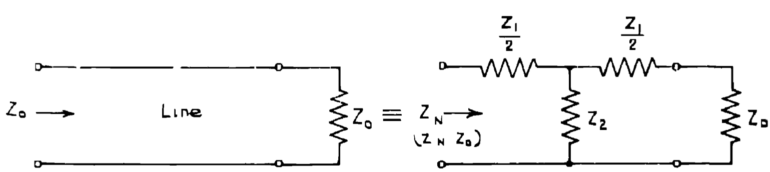


FIG. 728.—(a) Short line terminated in Z_0—Equivalent T section

line is terminated in Z_0 it will behave as an infinite line and have an input impedance Z_0. Since the equivalent T section represents the line, it also must have an input impedance Z_0 when it is terminated in Z_0.

Let the equivalent T section have series arms $\frac{Z_1}{2}$, $\frac{Z_1}{2}$, and shunt arm Z_2 (Fig. 728a).

Then $$Z_{IN} = \frac{Z_1}{2} + \frac{Z_2\left(\frac{Z_1}{2} + Z_0\right)}{\frac{Z_1}{2} + Z_2 + Z_0}$$

But $$Z_{IN} = Z_0$$

$$\therefore \quad Z_0^2 = \frac{Z_1^2}{4} + Z_1 Z_2 \tag{1}$$

or
$$Z_0 = \sqrt{\frac{Z_1^2}{4} + Z_1 Z_2} \tag{2}$$

Thus Z_0 for the T section, and hence for the line, may be determined if Z_1 and Z_2 can be found. This will require two equations which may be obtained by measuring the input impedance using two different terminating impedances. For convenience these terminations will be taken as infinity and zero.

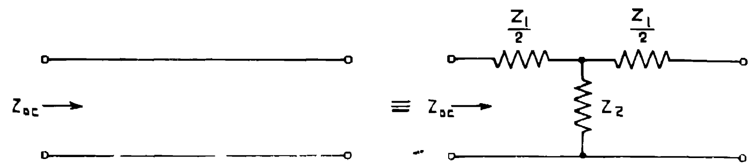


FIG. 728.—(*b*) Short line on open circuit—Equivalent T section on open circuit.

Let the input impedance with an infinite-impedance termination, *i.e.* open-circuit, be Z_{oc}.

Considering the equivalent T section on open-circuit (*see* Fig. 728*b*) :—

$$Z_{oc} = \frac{Z_1}{2} + Z_2 \tag{3}$$

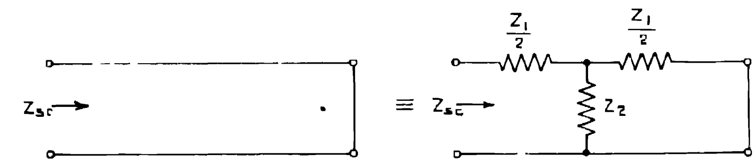


FIG. 728.—(*c*) Short line on short-circuit—Equivalent T section on short circuit.

Let the input impedance with a zero-impedance termination, *i.e.*, short-circuit, be Z_{sc}.

Considering the equivalent T section on short-circuit (*see* Fig. 728*c*) :—

$$Z_{sc} = \frac{Z_1}{2} + \frac{\frac{Z_1 Z_2}{2}}{\frac{Z_1}{2} + Z_2}$$

i.e.
$$Z_{sc} = \frac{\frac{Z_1^2}{4} + Z_1 Z_2}{\frac{Z_1}{2} + Z_2} \tag{4}$$

Equations 3 and 4 give two simultaneous equations from which Z_1 and Z_2 may be determined. At this stage, however, only Z_0 is

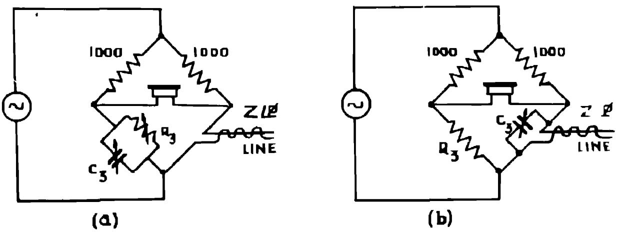
required, and this may be obtained directly by substituting equations 1 and 3 in equation 4 :—

$$Z_{sc} = \frac{Z_0^2}{Z_{oc}}$$

i.e.
$$Z_0 = \sqrt{Z_{oc} \cdot Z_{sc}} \qquad (5)$$

The characteristic impedance of a line is therefore the geometric mean of the open- and short-circuit impedances.

This gives a very convenient method for the determination of the characteristic impedance of a practical line, since Z_{oc} and Z_{sc} may readily be determined using an impedance bridge. The impedance bridges most frequently employed for such measurements



$$Z = \frac{R_3}{\sqrt{1+\omega^2 C_3^2 R_3^2}} \qquad Z = \frac{R_3}{\sqrt{1+\omega^2 C_3^2 R_3^2}}$$

$$\angle\phi = -\tan^{-1}\omega C_3 R_3 \qquad \angle\phi = \tan^{-1}\omega C_3 R_3$$

FIG. 729.—(*a*) Negative parallel bridge.
(*b*) Parallel resonance bridge.

are the inverted Wien (negative parallel) bridge (*see* Fig. 729*a*), for capacitative impedances, and the parallel resonance (negative parallel condenser-to-line) bridge (*see* Fig. 729*b*) for inductive impedances (*see* Chapter 5).

Example.—

The following measurements have been made on a line at 1600 c/s :—

$$Z_{oc} = 900\,\Omega \angle -30^\circ$$
$$Z_{sc} = 400\,\Omega \angle -10^\circ$$

What is the characteristic impedance of the line at 1600 c/s ?

$$\begin{aligned} Z_0 &= \sqrt{Z_{oc} \cdot Z_{sc}} \\ &= \sqrt{[900 \angle -30^\circ][400 \angle -10^\circ]} \\ &= \sqrt{360{,}000 \angle -40^\circ} \\ &= 600 \angle -20^\circ \quad \textit{Ans.} \end{aligned}$$

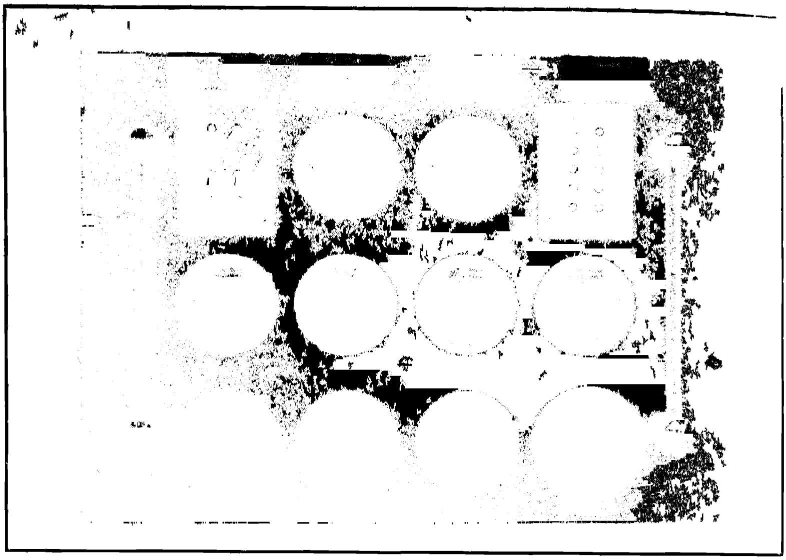


PLATE 35.—View of panel of typical impedance bridge.

CURRENTS AND VOLTAGES ALONG AN INFINITE LINE

Consider a current I_s applied to the sending end A of an infinite line (or a line terminated in Z_0) as in Fig. 730a. At the point B at a distance of one mile down the line let the current be I_1.

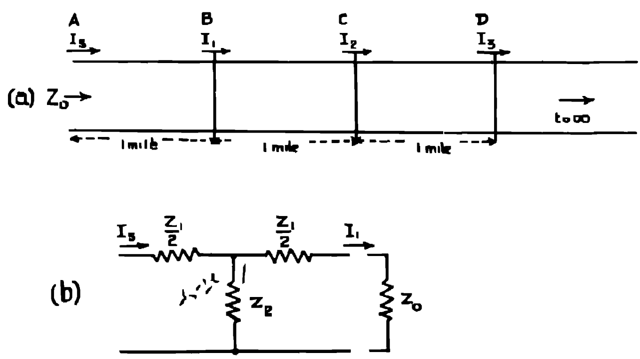


FIG. 730.—Currents along an infinite line.

Due to the loss introduced by the line, the current I_1 will be less than the current I_s; and since, in addition, a phase-shift will be introduced, the ratio $\frac{I_s}{I_1}$ will be a vector quantity.

A convenient way of representing a vector quantity is in the form e^{γ}, where γ is a complex quantity.

$$\text{Hence let } \frac{I_S}{I_1} = e^{\gamma}$$

γ is known as the "propagation constant" per mile of the line.

Considering the equivalent T section for the first mile of the line, γ may be determined in terms of the arms $\frac{Z_1}{2}$ and Z_2. From Fig. 730*b*, it will be seen that :—

$$I_1 = \frac{Z_2}{Z_2 + \frac{Z_1}{2} + Z_0} \cdot I_S$$

$$\text{Hence } e^{\gamma} = \frac{I_S}{I_1} = 1 + \frac{Z_1}{2Z_2} + \frac{Z_0}{Z_2} \tag{6}$$

$$\text{giving} \qquad \gamma = \log_e \left[1 + \frac{Z_1}{2Z_2} + \frac{Z_0}{Z_2}\right] \tag{7}$$

At a distance two miles down the line, at point C, let the current be I_2. Since the section of the line between B and C is identical with that between A and B, it follows that it may be represented by the same equivalent T section.

$$\text{Thus} \qquad \frac{I_1}{I_2} = 1 + \frac{Z_1}{2Z_2} + \frac{Z_0}{Z_2} = e^{\gamma}$$

Similarly it will be seen that, for the section between C and D :—

$$\frac{I_2}{I_3} = e^{\gamma}$$

where I_3 is the current at D, three miles down the line For the n^{th} section :—

$$\frac{I_{n-1}}{I_n} = e^{\gamma}$$

where I_{n-1} and I_n are the currents at distances $(n - 1)$ and n miles down the line respectively.

$$\text{Now} \qquad \frac{I_S}{I_1} = e^{\gamma}$$

$$\frac{I_S}{I_2} = \frac{I_S}{I_1} \cdot \frac{I_1}{I_2} = e^{2\gamma}$$

$$\frac{I_S}{I_3} = \frac{I_S}{I_1} \cdot \frac{I_1}{I_2} \cdot \frac{I_2}{I_3} = e^{3\gamma}$$

and, in the general case :—

$$\frac{I_S}{I_n} = \frac{I_S}{I_1} \cdot \frac{I_1}{I_2} \cdot \cdot \cdot \cdot \frac{I_{n-1}}{I_n} = e^{n\gamma}$$

From this, it follows that:—

$$I_n = I_S \cdot e^{-n\gamma} \tag{8}$$

This is the general equation for the current at a point distant n miles down an infinite line, in terms of the current at the sending end I_S, and the propagation constant per mile. It applies for any value of n.

A similar equation can be derived for voltage, since at all points along an infinite line the ratio of voltage to current is equal to the characteristic impedance Z_0.

Thus $$\frac{E_S}{I_S} = \frac{E_1}{I_1} = \frac{E_2}{I_2} = \frac{E_3}{I_3} = \ldots\ldots = \frac{E_n}{I_n} = Z_0$$

Hence $$\frac{E_S}{E_n} = \frac{I_S}{I_n}$$

But $$\frac{I_S}{I_n} = e^{n\gamma}$$

Therefore $$\frac{E_S}{E_n} = e^{n\gamma}$$

from which it follows that :—

$$E_n = E_S \cdot e^{-n\gamma} \tag{9}$$

Attenuation and phase constants

The propagation constant γ is a complex quantity; let it be equal to $\alpha + j\beta$.

Thus, for one mile of line :—

$$\frac{I_S}{I_1} = e^{\gamma} = e^{\alpha + j\beta}$$

$$= e^{\alpha} \underline{/\beta}$$

Hence $\left|\frac{I_S}{I_1}\right| = e^{\alpha}$, and the angle of $\frac{I_S}{I_1}$ is $\underline{/\beta}$. It follows that $\alpha = \log_e \left|\frac{I_S}{I_1}\right|$. Note that e^{α} gives the ratio of the absolute value of the current sent, to that of the current received, while β gives the phase angle between the two currents.

α is known as the attenuation constant per mile of the line, and is measured in nepers per mile.

β is known as the phase constant or wavelength constant per mile of the line, and is measured in radians per mile.

If the length of the line is n miles :—

$$\frac{I_S}{I_n} = e^{n\gamma} = e^{n\alpha + jn\beta}$$

$$= e^{n\alpha} \underline{/n\beta}$$

The attenuation of such a line is thus $n\alpha$ nepers, and the phase-shift is $n\beta$ radians.

It must be noted that, throughout this chapter, the values obtained for attenuation and phase-shift are in nepers and radians. These results can be converted into the more convenient units for practical work, the decibel and the degree, by multiplying by 8·686 and 57·3 respectively, or by using the Conversion Tables on pages 800, 804, 838 and 839.

Conclusions.—

In the general case of an infinite line or a short line terminated in its characteristic impedance and having a propagation constant γ, the current I at any point distant x from the sending end will be given by :—

$$I = I_S \, . \, e^{-\gamma x} \tag{10}$$

$$= I_S \, e^{-\alpha x} \underline{/ -\beta x} \tag{11}$$

where I_S is the sending-end current.

The voltage E at any point distant x from the sending end will be given by :—

$$E = E_S \, . \, e^{-\gamma x} \tag{12}$$

$$= E_S \, e^{-\alpha x} \underline{/ -\beta x} \tag{13}$$

where E_S is the sending-end voltage.

Example.—

A twisted D8 cable has, at 1600 c/s, an attenuation of 3·0 db per mile and a phase constant of 0·319 radians per mile in dry weather. If 2 volts at 1600 c/s are applied to the sending end, what will be the voltage at a point 10 miles down the line when the line is correctly terminated (*i.e.* terminated in its characteristic impedance) ?

The voltage at a point distant x miles from the sending end is :—

$$E = E_S \, . \, e^{-\gamma x}$$

$$= E_S \, . \, e^{-\alpha x} \underline{/ -\beta x}$$

where α is in nepers per mile
and β is in radians per mile.

In this case, attenuation at 3·0 db/mile is equivalent to $3 \cdot 0 \times 0 \cdot 115$ nepers per mile, *i.e.* 0·345 nepers/mile.

The required voltage $E = 2 \, . \, e^{-0 \cdot 345 \times 10}, \underline{/ - 0 \cdot 319 \times 10}$

$= 2 \, . \, e^{-3 \cdot 45}, \underline{/ - 3 \cdot 19}$

$= 0 \cdot 0635$ volts, $\underline{/ - 3 \cdot 19}$. *Ans.*

Thus the voltage at a point 10 miles from the sending end is 0·0635 volts, lagging 3·19 radians, *i.e.* 182° 47′, behind the sending end voltage.

Graphical representation of current and voltage distribution along an infinitely long line

Current.

It has been seen that the current I at any point distant x from the sending end of a line is given by :—

$$I = I_S \,.\, e^{-\alpha x} \underline{/- \beta x}$$

To represent this equation graphically would require a three dimensional figure, because I_S itself is an alternating current and therefore I varies both with the distance x and with the time t

If I_S has a frequency f, it may be represented by a rotating vector of length $I_{S max}$ rotating with angular velocity ω radians per second, where $\omega = 2\pi f$. The projection of this vector on to the vertical axis will give the instantaneous value of I_S (*see* Fig 731)

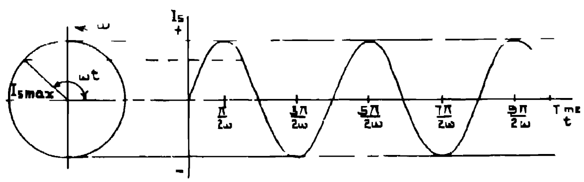


FIG 731 —Rotating vector and graphical representation of I_S against time

At a point distant x down the line, the current I will have the same frequency f, but, due to the attenuation of the line, the maximum amplitude will have been reduced from $I_{S max}$ to $I_{S max}\, e^{-\alpha x}$. The current I will therefore vary sinusoidally between the limit $I_{S max}\, e^{-\alpha x}$ at the peak of the positive half-cycle and $-\, I_{S max}.e^{-\alpha x}$ at the peak of the negative half-cycle (*see* Fig. 732)

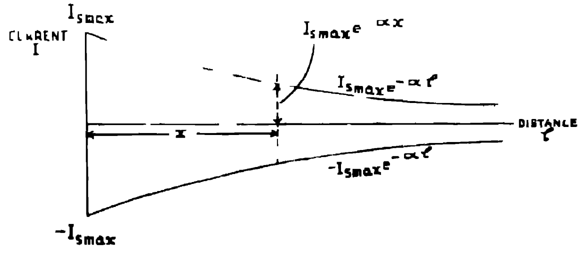


FIG 732 —Decrease in amplitude of current along an infinite line

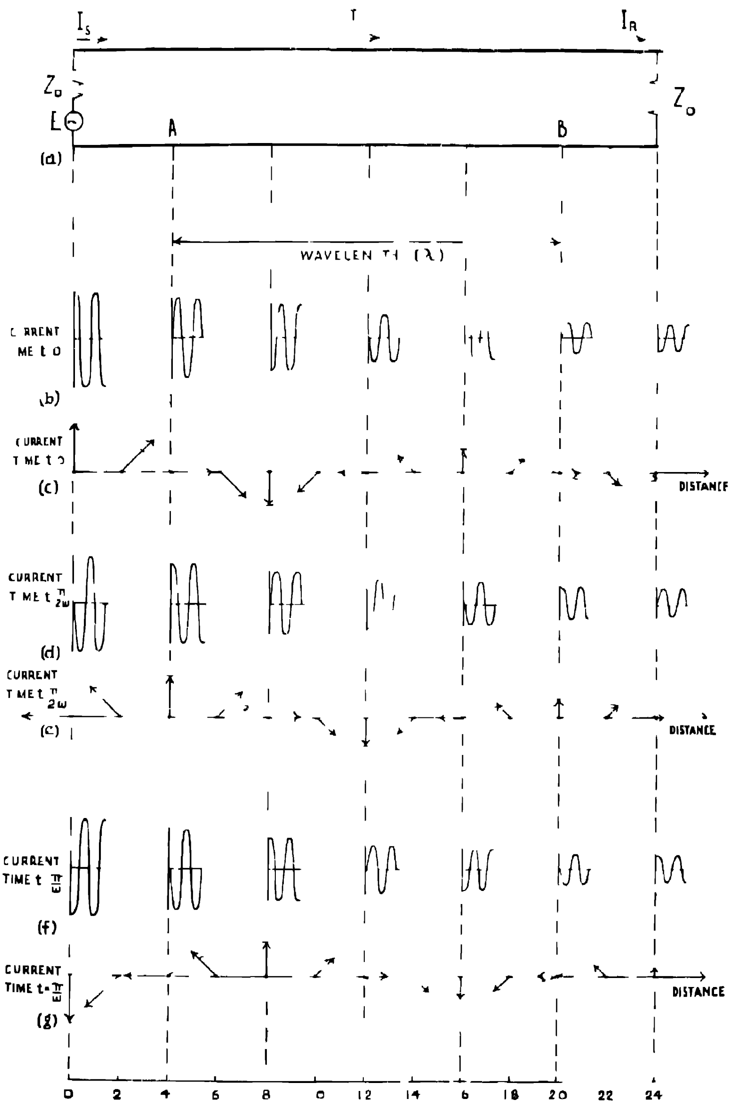


FIG. 733 —Magnitude and phase of current along an infinite line.

Due to the phase constant β, the current I will lag by a phase angle βx on the current I_s.

Both the magnitude and the phase of the current I at various points along the line (*see* Fig. 733*a*) may be seen from the sine waves in Fig. 733*b* ; each of these represents the variation of current with time at the point on the line corresponding to the vertical line at the left of the sine wave.

In place of these sine waves, however, it is more convenient to consider their corresponding rotating vectors, as shown in Fig. 733 These vectors are drawn to correspond with the instant at which I_s is at its positive peak ; but it must be remembered that all these vectors are rotating in an anti-clockwise direction with angular velocity ω. Thus at a time $\frac{\pi}{2\omega}$ seconds later, I_s will be at zero and approaching its negative peak, and the corresponding sine wave and vector diagrams will be as shown in Figs. 733*d* and *e* respectively ; while at a time $\frac{\pi}{\omega}$ seconds, I_s will be at its negative peak, and the corresponding sine wave and vector diagrams as in Figs. 733*f* and *g*.

Alternatively, if one instant of time be considered, a picture may be obtained of the current distribution along the line at that instant. Thus, corresponding to the vector diagram in Fig. 733*c* which shows I_s at its positive peak, one can draw the dotted curve along the line. Similarly, the dotted curves of Figs. 733*e* and *g* show the current distribution at all points along the line at the instants when I_s is at zero and at its negative peak respectively.

Voltage.

The voltage at any point distance x from the sending end is given by :-

$$E = E_s \,.\, e^{-\alpha x} \underline{/- \beta x}$$

The ratio of voltage to current at all points down an infinite line is equal to Z_0, the characteristic impedance of the line.

Thus if $Z_0 = |Z_0| \underline{/- \varphi}$, the voltage at any point will lag on the current by the angle φ. Bearing in mind this difference in phase, the voltage may now be represented in a similar manner to the current.

Fig. 734*b* and *c* show the instantaneous distributions of current and voltage respectively along a line (*see* Fig. 734*a*) having a characteristic impedance $1000 \underline{/- 45^\circ}$.

The full curve in Fig. 734*d* shows the distribution of power $P = |E|\,.\,|I| \cos\varphi$ along the line, while the broken line shows the same power distribution expressed in decibels referred to 1mW.

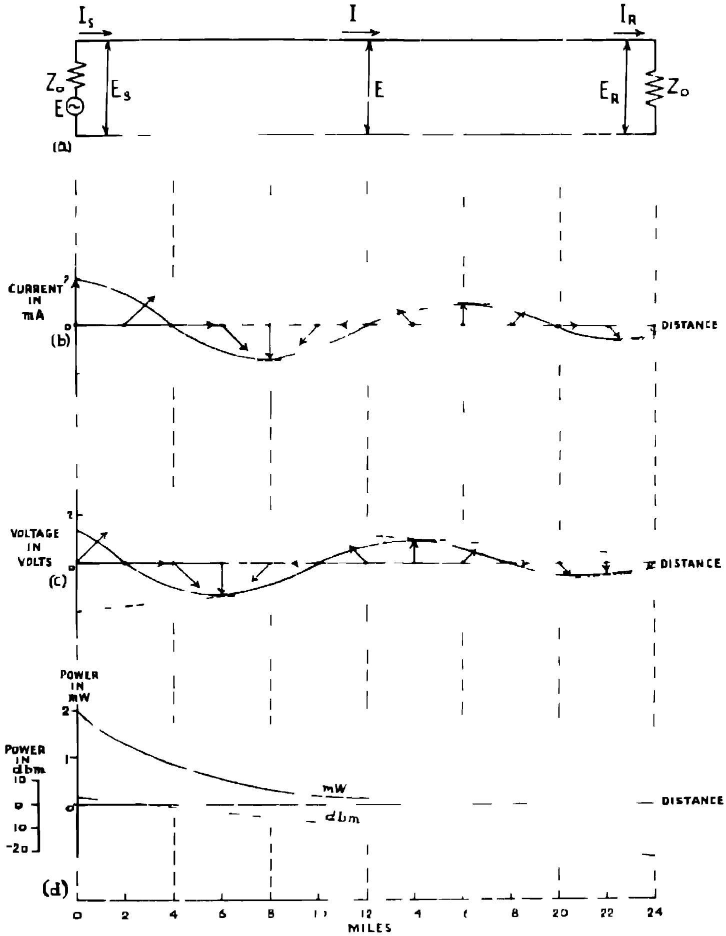


FIG 734 —Distribution of current voltage and power along a correctly terminated line

Logarithmic spiral representation of current and voltage distribution along an infinite line

The diagrams dealt with so far refer to one instant of time only, and in order to obtain a complete picture of the current or voltage distribution at all instants one diagram would be required for each position of the I_S or E_S vectors

Just as a waveform that varies in amplitude with time may be represented by a rotating vector so may a waveform that varies in amplitude with distance Thus the instantaneous picture given by the dotted curve in Fig 733c may be represented by a vector of length $I_{Smax}\ \epsilon^{-\alpha x}$ and angle βx as shown in Fig 735 As x increases, this vector rotates in a clockwise direction, but at the same time its modulus decreases logarithmically and the locus of the vector is, in fact a logarithmic spiral The projection of

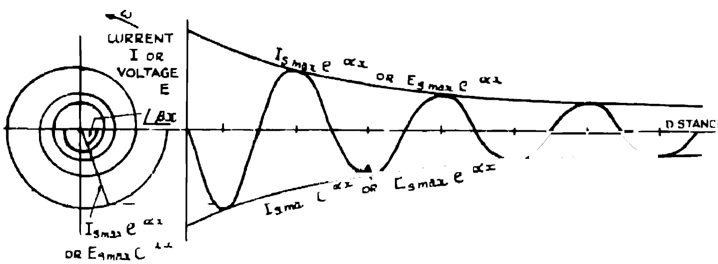


FIG 735 Logarithmic spiral representation of current and voltage at any point along a line and at any time

the rotating vector on to the vertical axis will give the instantaneous value of current at the point down the line corresponding to the distance x

If the complete logarithmic spiral is now rotated in an anti clockwise direction with angular velocity ω radians per second and the projection taken the value of current or voltage will be obtained for all distances (x) down the line and for all times (t)

Wavelength and velocity

The wavelength λ is the distance between any point and the next point along the line at which the current (or voltage) is in the same phase In Fig 733 A and B are two such points Although the current at A reaches a maximum at the same instant as the current at B, the current at A is really leading by 2π radians on the current at B The phase-shift along the line is β radians per mile. Hence the distance λ must be $\frac{2\pi}{\beta}$ miles, that is :—

$$\lambda = \frac{2\pi}{\beta}$$

The velocity* of propagation (v) = frequency × wavelength

Thus $$v = \frac{2\pi f}{\beta}$$

or $$v = \frac{\omega}{\beta} \tag{15}$$

Example:—

At 1600 c/s, ω = 10,000 radians/second.

For an airline β = 0.055 radians per mile

$$\therefore \quad \lambda = \frac{2\pi}{0.055} = 114.2 \text{ miles}$$

$$\text{and} \quad v = \frac{10{,}000}{0.055} = 182{,}000 \text{ miles/second.}$$

Therefore the time taken for a wave of this frequency to travel 100 miles is 0.55 milliseconds.

For a loaded underground cable β = 1.0 radian per mile at 1600 c/s

$\therefore$ λ = 6.28 miles

and v = 10,000 miles/second.

In this case the time taken for a signal to travel 100 miles is 10 milliseconds.

LINE CONSTANTS

It has already been seen that a practical line has a characteristic impedance Z_0, a propagation constant P, an attenuation constant α, and a phase constant β. These are known as the "secondary line constants." Although they are referred to as constants it should be noted that, in general, all will vary if the frequency is changed.

The "primary line constants" (which for the purpose of transmission theory, are assumed to be independent of frequency) are R, G, L, and C, where:—

R is the resistance per mile of the line measured in ohms,

G is the leakance per mile of the line measured in mhos,

L is the inductance per mile of the line measured in henries,

C is the capacitance per mile of the line measured in farads.

They are measured considering both conductors, *i.e.* per mile loop.

These primary constants may be obtained by measurements on a sample of the line.

---

* This is the phase or wave velocity. The group velocity (*i.e.*, the velocity at which the energy is transferred along the line) is $\frac{d\omega}{d\beta}$.

Relationship between primary and secondary line constants

Consider a short length of line l miles long. This short section will have a resistance Rl, a leakance Gl, an inductance Ll, and a capacitance Cl.

Its characteristic impedance will be Z_0, the same as that of the complete line. Its propagation constant will be γl, where γ is the propagation constant per mile of the complete line.

This short section of line may be represented by a T network as shown in Fig. 736b.

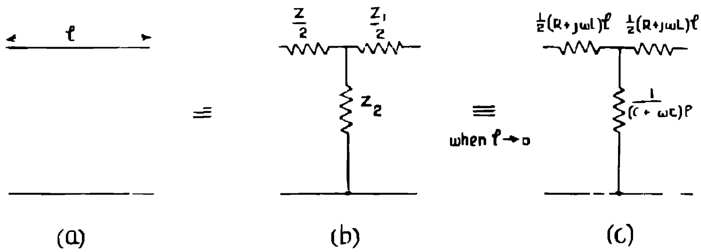


FIG. 736. T sections equivalent to short length of line

If the length of the section is very small Z_1 will be approximately equal to the series impedance of the section, i.e. to $Rl + j\omega Ll$, and Z_2 will be approximately equal to the shunt impedance of the section, i.e. $\dfrac{1}{Gl + j\omega Cl}$.

The accuracy of this statement increases as l decreases, and in order to obtain an accurate answer it will be assumed that the section is so small that l tends to zero.

Determination of Z_0

It has been shown that for a T section

$$Z_0 = \sqrt{\frac{Z_1^2}{4} + Z_1 Z_2} \qquad \text{(Equation 2)}$$

Hence in this case since $Z_1 = (R + j\omega L)l$ and $Z_2 = \dfrac{1}{(G + j\omega C)l}$

$$Z_0 = \sqrt{\frac{(R + j\omega L)^2 l^2}{4} + \frac{(R + j\omega L)\,l}{(G + j\omega C)\,l}}$$

$$= \sqrt{\frac{R + j\omega L}{G + j\omega C} + \frac{(R + j\omega L)^2}{4}\,l^2}$$

As $l \to 0$ terms containing l^2 may be neglected giving:—

$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \qquad (16)$$

This equation is very important, since it enables both the

modulus and the angle of Z_0 to be calculated from a knowledge of the primary line constants.

Example.—

A line has the following primary line constants :—

$R = 100$ ohms per mile.

$G = 1\cdot5 \times 10^{-6}$ mhos per mile.

$L = 0\cdot001$ henries per mile.

$C = 0\cdot062$ microfarads per mile.

Find the characteristic impedance in modulus-and-angle form at 1000 c/s.

$$
\begin{aligned}
R + j\omega L &= 100 + j \times 2\pi \times 1000 \times 0\cdot001 \\
&= 100 + j\,6\cdot283 \\
&= 100\cdot2 \,\angle\, \underline{3^\circ\ 36'} \\
G + j\omega C &= 1\cdot5 \times 10^{-6} + j \times 2\pi \times 1000 \times 0\cdot062 \times 10^{-6} \\
&= (1\cdot5 + j\,389\cdot5) \times 10^{-6} \\
&= 389\cdot5 \times 10^{-6} \,\angle\, \underline{89^\circ\ 48'} \\
Z_0 &= \sqrt{\frac{R + j\omega L}{G + j\omega C}} \\
&= \sqrt{\frac{100\cdot2 \,\angle\, \underline{3^\circ\ 36'}}{389\cdot5 \times 10^{-6} \,\angle\, \underline{89^\circ\ 48'}}} \\
&= 507 \,\angle\, \underline{-43^\circ\ 6'} \quad \textit{Ans}
\end{aligned}
$$

Equation 16 can be expressed in the modulus and angle form :—

$$
\begin{aligned}
Z_0 &= \sqrt{\frac{R + j\omega L}{G + j\omega C}} \\
&= \left[\frac{R + j\omega L}{G + j\omega C}\right]^{\frac{1}{2}} \\
&= \left[\frac{\sqrt{R^2 + \omega^2 L^2},\ \tan^{-1}\dfrac{\omega L}{R}}{\sqrt{G^2 + \omega^2 C^2},\ \tan^{-1}\dfrac{\omega C}{G}}\right]^{\frac{1}{2}} \\
&= \left[\frac{\sqrt{R^2 + \omega^2 L^2}}{\sqrt{G^2 + \omega^2 C^2}},\ \tan^{-1}\frac{\omega L}{R} - \tan^{-1}\frac{\omega C}{G}\right]^{\frac{1}{2}} \\
&= \sqrt[4]{\frac{R^2 + \omega^2 L^2}{G^2 + \omega^2 C^2}},\ \tfrac{1}{2}\left(\tan^{-1}\frac{\omega L}{R} - \tan^{-1}\frac{\omega C}{G}\right) \qquad (17)
\end{aligned}
$$

Alternatively, since $Z_0 = \sqrt{\dfrac{R + j\omega L}{G + j\omega C}}$

it follows that $Z_0 = \sqrt{\dfrac{\omega L - jR}{\omega C - jG}}$ $\left(\text{multiplying by} \sqrt{\dfrac{-j}{-j}}\right)$

Hence, instead of equation 17, one can write —

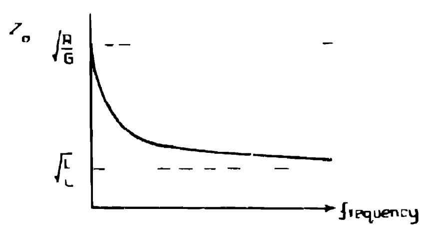
$$Z_0 = \sqrt[4]{\frac{R^2 + \omega^2 L^2}{G^2 + \omega^2 C^2}} \angle \frac{1}{2}\left[\tan^{-1}\frac{G}{\omega C} - \tan^{-1}\frac{R}{\omega L}\right] \qquad (18$$

It will be seen from equation 17 that —

When ω is very small, $|Z_0| \rightarrow \sqrt{\dfrac{R}{G}}$

When ω is very large, $|Z_0| \rightarrow \sqrt{\dfrac{L}{C}}$

Since $\dfrac{R}{G}$ is in all cases greater than $\dfrac{L}{C}$, the variation of Z_0 with frequency expected for a practical line will be as in Fig 737.



Variation of Z_0 with frequency

Fig 737 — Characteristic impedance of a line

Taking 70 lb Cd Cu multi airline as an example of an open-wire line, Fig 738 shows how $|Z_0|$ and $\underline{/\varphi_0}$ vary with frequency over the audio and carrier range for a practical line.

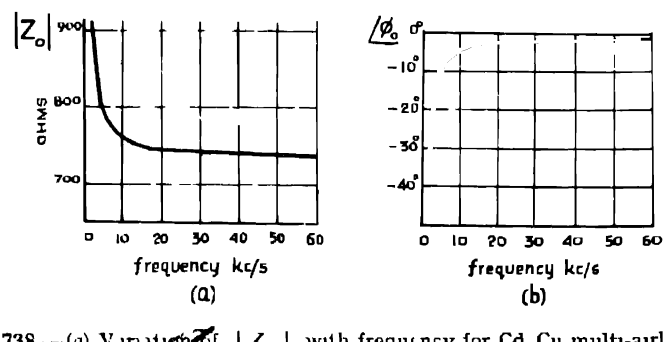


Fig 738 —(a) Variation of $|Z_0|$ with frequency for Cd–Cu multi-airline
(b) Variation of $\underline{/\varphi_0}$ with frequency for Cd–Cu multi airline

Determination of γ

For the T section of Fig 736*b*, the propagation constant is γl. Since the series and shunt arms of the T section are $\frac{Z_1}{2}$ and Z_2 respectively, it follows that

$$e^{\gamma l} = 1 + \frac{Z_1}{2Z_2} + \frac{Z_0}{Z_2}$$

But as $l \rightarrow 0$, $Z_1 = (R + j\omega L)\, l$

$$\text{and } Z_2 = \frac{1}{(G + j\omega C)\, l}$$

Hence
$$e^{\gamma l} = 1 + \frac{(R + j\omega L)(G + j\omega C)\, l^2}{2} + Z_0 (G + j\omega C)\, l$$

$$= 1 + \frac{(R + j\omega L)(G + j\omega C)\, l^2}{2} + \sqrt{\frac{R + j\omega L}{G + j\omega C}}\,(G + j\omega C)\, l$$

$$= 1 + \sqrt{(R + j\omega L)(G + j\omega C)}\, l + \frac{(R + j\omega L)(G + j\omega C)\, l^2}{2} \qquad (19)$$

But by the exponential series

$$e^{\gamma l} = 1 + \gamma l + \frac{\gamma^2 l^2}{\underline{|2}} + \ldots \text{ to } \infty$$

As $l \rightarrow 0$, terms containing l^3 and higher may be neglected,

$$e^{\gamma l} = 1 + \gamma l + \frac{\gamma^2 l^2}{2} \qquad (20)$$

Comparing equations 19 and 20 it is seen that

$$\gamma = \sqrt{(R + j\omega L)(G + j\omega C)} \qquad (21)$$

This equation is very important since it enables γ to be calculated from a knowledge of the primary line constants.

Since $\gamma = \alpha + j\beta$,

$$\alpha + j\beta = \sqrt{(R + j\omega L)(G + j\omega C)} \qquad (22)$$

If γ is to be calculated equation 21 should be used, and the result obtained in rectangular notation, then the real part will be the attenuation constant α and the imaginary part the phase constant β.

If γ is obtained in the modulus-and angle notation, say —

$$\gamma = P \angle \theta \qquad (23)$$

the attenuation and phase constants may be found from the equations —

$$\alpha = P \cos \theta \qquad (24)$$

$$\beta = P \sin \theta \qquad (25)$$

The primary and secondary line constants for various types of line are given in Table XX.

Example.—

A sample of field quad cable has the following primary line constants :—

$R = 78$ ohms per mile loop.
$G = 62$ micromhos per mile
$L = 1 \cdot 75$ millihenries per mile loop.
$C = 0 \cdot 0945$ microfarads per mile.

Required to calculate at 1600 c/s ($\omega \simeq 10{,}000$ radians second) the following :-

(i) Characteristic impedance Z_0.
(ii) Attenuation constant (α), in nepers and decibels per mile.
(iii) Phase constant (β), in radians and degrees per mile.
(iv) Wavelength (λ), in miles.
(v) Velocity (v) in miles per second.
(vi) Time for a wave to travel 100 miles along the line, in milliseconds.

$$R + j\omega L = 78 + j10{,}000 \times 1 \cdot 75 \times 10^{-3}$$
$$= 78 + j17 \cdot 5$$
$$= 79 \cdot 94 \;\angle 12^\circ\, 39'$$
$$G + j\omega C = 10^{-6} \times 62 + j10{,}000 \times 0 \cdot 0945 \times 10^{-6}$$
$$= 10^{-6}\,(62 + j945)$$
$$= 10^{-6} \times 947 \;\angle 86^\circ\, 15'$$
$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}} = \sqrt{\frac{79 \cdot 94}{947 \times 10^{-6}}}, \;\angle -36^\circ\, 48'$$

i.e., $Z_0 = 290, \angle -36^\circ\, 48'$ *Ans.* (*i*)

$$\gamma = \sqrt{(R + j\omega L)(G + j\omega C)} = \sqrt{79 \cdot 94 \times 10^{-6} \times 947}, \;\angle 49^\circ\, 27'$$

i.e., $\gamma = 0 \cdot 275, \angle 49^\circ\, 27'$

i.e., $\gamma = 0 \cdot 179 + j0 \cdot 209$

$\alpha = 0 \cdot 179$ nepers per mile $= 1 \cdot 56$ db per mile. *Ans.* (*ii*).

$\beta = 0 \cdot 209$ radians per mile $= 11^\circ\, 58'$ per mile. *Ans.* (*iii*).

The wavelength $\lambda = \dfrac{2\pi}{\beta} = \dfrac{2\pi}{0 \cdot 209} = 30$ miles. *Ans.* (*iv*).

The velocity $v = \dfrac{\omega}{\beta} = \dfrac{10{,}000}{0 \cdot 209} = 47{,}840$ miles per second. *Ans.* (*v*)

The time taken for a wave to travel 100 miles is :—

$$t = \frac{100}{47{,}840} \text{ seconds}$$
$$= 0 \cdot 00209 \text{ seconds}$$
$$= 2 \cdot 09 \text{ milliseconds. } \textit{Ans. (vi).}$$

TABLE XX

PRIMARY and SECONDARY LINE CONSTANTS (at 1600 c s)

Type of l	I l	C t f t h	I	I H	μh	C I F r	I/L	—	Z l l	φ gi	t r	H r l l r l	r le	miles per sec
300 lb Cu Air	l 1 1		[illegible]			[illegible]	1 4	11		4 [illegible]	[illegible]	[illegible]	11	179 000
200 lb C Air ne	l 1 1		[illegible]	[illegible]	1	[illegible]	4	11		[illegible] 4		[illegible]	111	177 000
150 lb C Air	l		11			[illegible]	1	1 [illegible]	[illegible]	[illegible]			1	17[illegible] 000
150 lb C I C Air	l					[illegible]		1 9	9	1	9	[illegible]	11	[illegible] 000
70 lb C I C Air e			4			[illegible]	4	4	4	1[illegible]	l	1	1 [illegible]	164 0[illegible]0
40 lb Cd Cu Air				4 [illegible]		9	4		9		4		[illegible]	1[illegible] 000
40 lb I C O I I	l		4	1					4	[illegible]	13	14	[illegible]	74 [illegible]
40 lb I C O I l	[illegible] H 1 1	9[illegible]	4 3			[illegible]		1 [illegible]	1	1	19		[illegible]	14 000
70 lb P C Q I I	l						[illegible]	1 [illegible]	3[illegible]	−14	1 44	[illegible]	[illegible] 4	59 500
20 lb I C Q l	[illegible] 1 1	3	91 3	[illegible]		[illegible]	11		11	3	3	[illegible]	[illegible]	14 000
F eld [illegible] ad ca l	l			[illegible]		94[illegible]	4 [illegible]	1	9	3 [illegible]	[illegible]	9	4	46 000
DS [illegible] tl sp	l		1[illegible]			0 [illegible]	9	1 [illegible]	1	9	4	111	[illegible]	90 000
DS calle t t	l nl		1[illegible]		9[illegible]	[illegible]		0 1	4[illegible]	3	4	6 34	1[illegible] 4	29 000
DS cable t s d s	l		19	1	3[illegible]	[illegible]	[illegible]	[illegible]	9[illegible]	−30	4	[illegible]	1 [illegible]	18 000

Evaluation of α and β

Equation 21 may be converted into the modulus-and-angle notation, in which case :—

$$\gamma = \sqrt[4]{(R^2 + \omega^2 L^2)(G^2 + \omega^2 C^2)}, \tfrac{1}{2}\left[\tan^{-1}\frac{\omega L}{R} + \tan^{-1}\frac{\omega C}{G}\right] \tag{26}$$

$$\text{Hence } |\gamma| = \sqrt[4]{(R^2 + \omega^2 L^2)(G^2 + \omega^2 C^2)}$$

$$\text{But } \gamma = \alpha + j\beta$$

$$\therefore \quad |\gamma| = \sqrt{\alpha^2 + \beta^2}$$

Thus $\sqrt{\alpha^2 + \beta^2} = \sqrt[4]{(R^2 + \omega^2 L^2)(G^2 + \omega^2 C^2)}$

Squaring both sides :

$$\alpha^2 + \beta^2 = \sqrt{(R^2 + \omega^2 L^2)(G^2 + \omega^2 C^2)} \tag{27}$$

From equation 22 :—

$$\alpha + j\beta = \sqrt{(R + j\omega L)(G + j\omega C)}$$

Squaring both sides :—

$$(\alpha + j\beta)^2 = (R + j\omega L)(G + j\omega C)$$

$$\alpha^2 + 2j\alpha\beta - \beta^2 = RG + j\omega LG + j\omega CR - \omega^2 LC$$

Equating real terms on both sides :—

$$\alpha^2 - \beta^2 = RG - \omega^2 LC \tag{28}$$

Adding equations 28 and 27 :—

$$2\alpha^2 = \sqrt{(R^2 + \omega^2 L^2)(G^2 + \omega^2 C^2)} + (RG - \omega^2 LC)$$

$$\therefore \alpha = \sqrt{\tfrac{1}{2}\left[\sqrt{(R^2 + \omega^2 L^2)(G^2 + \omega^2 C^2)} + (RG - \omega^2 LC)\right]} \tag{29}$$

Subtracting 28 from 27 :—

$$2\beta^2 = \sqrt{(R^2 + \omega^2 L^2)(G^2 + \omega^2 C^2)} - (RG - \omega^2 LC)$$

$$\therefore \beta = \sqrt{\tfrac{1}{2}\left[\sqrt{(R^2 + \omega^2 L^2)(G^2 + \omega^2 C^2)} - (RG - \omega^2 LC)\right]} \tag{30}$$

Practical formulae for Z_0, α, and β for unloaded cables

In practice, certain approximations are made to obtain simplified expressions for Z_0, α and β. In the case of an underground cable, small diameter conductors are used in order to obtain the maximum number of conductors for a given overall diameter of the cable. An underground cable will, therefore, in general, have a fairly large resistance R per mile ; and, due to the small spacing between the conductors, a large capacitance C and a small inductance L per mile. The leakance G per mile is very small, due to the good insulation between conductors in a well-laid and well-maintained cable.

To take a particular case, the constants for an air-spaced paper-insulated 20 lb. underground cable are given as $R = 88$ ohms per

mile, $G = 10^{-6}$ mhos per mile, $L = 0{\cdot}001$ henries per mile and $C = 0{\cdot}065\mu F$ per mile. Thus, taking the extreme limits of the audio frequency band as 200 and 3200 c/s, it may be seen that in this band ωL is always less than 20 and ωC is always greater than 80×10^{-6}

The permissible approximations for unloaded underground cables at audio frequencies are therefore that $\omega C \gg G$, and that $\omega L \ll R$. These approximations will clearly give more accurate results at the lower than at the higher audio frequencies

Characteristic impedance Z_0

$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \quad \text{(equation 16)}$$

When $R \gg \omega L$ and $\omega C \gg G$

$$Z_0 \simeq \sqrt{\frac{R}{j\omega C}} = \sqrt{\frac{R}{\omega C}} \angle \tfrac{1}{2}(0 - 90^\circ)$$

$$\text{i.e., } Z_0 \simeq \sqrt{\frac{R}{\omega C}} \angle -45^\circ \qquad (31)$$

Propagation constant γ

$$\gamma = \sqrt{(R + j\omega L)(G + j\omega C)} \quad \text{(equation 21)}$$

When $R \gg \omega L$ and $\omega C \gg G$

$$\gamma \simeq \sqrt{j\omega CR} = \sqrt{\omega CR}\angle 45^\circ$$

$$= \sqrt{\omega CR}\cos 45^\circ + j\sqrt{\omega CR}\sin 45^\circ$$

$$= \sqrt{\frac{\omega CR}{2}} + j\sqrt{\frac{\omega CR}{2}}$$

But $\gamma = \alpha + j\beta$

$$\therefore \quad \alpha \simeq \sqrt{\frac{\omega CR}{2}} \text{ nepers per mile} \qquad (32)$$

$$\text{and } \beta \simeq \sqrt{\frac{\omega CR}{2}} \text{ radians per mile} \qquad (33)$$

Example—

An underground cable has the following constants —

$R = 44$ ohms per mile loop,

$G = 1$ micromho per mile,

$L = 0{\cdot}001$ henries per mile loop,

$C = 0{\cdot}065$ microfarads per mile.

Find the approximate values of Z_0, α and β at 400 c/s and 1600 c/s

$$Z_0 \simeq \sqrt{\frac{R}{\omega C}} \angle -45^\circ$$

At 400 c/s, $\omega = 2500$ radians per second,

$$\therefore Z_0 \simeq \sqrt{\frac{44}{2500 \times 0\cdot065 \times 10^{-6}}} \angle -45^\circ$$

$$\therefore Z_0 \simeq 520 \angle -45^\circ$$

[Accurately, $Z_0 = 521 \angle -43^\circ 14'$]

At 1600 c/s, $\omega = 10{,}000$ radians per second,

$$\therefore Z_0 \simeq \sqrt{\frac{44}{10\,000 \times 0\cdot065 \times 10^{-6}}} \angle -45^\circ$$

$$\therefore Z_0 \simeq 260 \angle -45^\circ$$

[Accurately, $Z_0 = 263 \angle -38^\circ 36'$]

It will be noted that the modulus of Z_0 for an unloaded underground cable approximates to a value that is inversely proportional to the square root of the frequency, whilst the angle approximates to $\angle -45^\circ$

$$\alpha = \beta = \sqrt{\frac{\omega CR}{2}}.$$

At 400 c/s,

$$\alpha = \beta = \sqrt{\frac{2\,500 \times 0\cdot065 \times 10^{-6} \times 44}{2}} = 0\cdot0598$$

Hence $\alpha = 0\cdot0598$ nepers per mile

$\beta = 0\cdot0598$ radians per mile

[Accurately, $\alpha = 0\cdot0582$ nepers per mile

and $\beta = 0\cdot0612$ radians per mile]

At 1600 c/s

$$\alpha = \beta = \sqrt{\frac{10{,}000 \times 0\cdot065 \times 10^{-6} \times 44}{2}} = 0\cdot120$$

Hence $\alpha = 0\cdot120$ nepers per mile

$\beta = 0\cdot120$ radians per mile

[Accurately, $\alpha = 0\cdot107$ nepers per mile

and $\beta = 0\cdot134$ radians per mile]

It will be noted that for an unloaded underground cable at audio frequencies, α and β approximate to values directly proportional to the square root of the frequency

LOADING OF LINES

Conditions for minimum attenuation

The attenuation constant α has been shown by equation 29 to be given by -

$$\alpha = \sqrt{\tfrac{1}{2}\left[\sqrt{(R^2 + \omega^2 L^2)(G^2 + \omega^2 C^2)} + (RG - \omega^2 LC)\right]}$$

It will be seen that α depends on the four primary line constants in addition to the frequency under consideration.

Value of L for minimum attenuation.—To determine the value of L for minimum attenuation when L only may be varied, differentiate α with respect to L, and equate to zero.

$$\frac{d\alpha}{dL} = \frac{1}{2}\frac{\frac{1}{2}\left[\frac{1}{2}\frac{2\,\omega^2 L\,(G^2+\omega^2 C^2)}{\sqrt{(R^2+\omega^2 L^2)\,(G^2+\omega^2 C^2)}} - \omega^2 C\right]}{\sqrt{\frac{1}{2}\left[\sqrt{(R^2+\omega^2 L^2)\,(G^2+\omega^2 C^2)} + (RG-\omega^2 LC)\right]}}$$

$$\therefore \frac{1}{2}\left[\frac{1}{2}\left\{\frac{2\,\omega^2 L\,(G^2+\omega^2 C^2)}{2\sqrt{(R^2+\omega^2 L^2)\,(G^2+\omega^2 C^2)}} - \omega^2 C\right\}\right] = 0.$$

Hence the condition for minimum attenuation is that .

$$\frac{\omega^2 L\,(G^2+\omega^2 C^2)}{\sqrt{(R^2+\omega^2 L^2)\,(G^2+\omega^2 C^2)}} - \omega^2 C = 0$$

i.e., $$\frac{L\,(G^2+\omega^2 C^2)}{\sqrt{(R^2+\omega^2 L^2)\,(G^2+\omega^2 C^2)}} = C$$

i.e. $$\frac{L\sqrt{G^2+\omega^2 C^2}}{\sqrt{R^2+\omega^2 L^2}} = C$$

i.e., $$L\sqrt{G^2+\omega^2 C^2} = C\sqrt{R^2+\omega^2 L^2}$$

Squaring both sides :—

$$L^2\,(G^2+\omega^2 C^2) = C^2\,(R^2+\omega^2 L^2)$$

i.e., $$L^2 G^2 + \omega^2 C^2 L^2 = C^2 R^2 + \omega^2 C^2 L^2$$

i.e., $$L^2 G^2 = C^2 R^2$$

i.e., $$L = \frac{CR}{G} \qquad (34)$$

Thus, if L is variable, the attenuation will be a minimum when :—

$$L = \frac{CR}{G}\text{ henries/mile}$$

This result is important because in practice L is normally less than this desired value, and hence the attenuation of a line can be reduced by artificially increasing L.

Value of C for minimum attenuation. In a similar manner, if C is considered as the only variable, its value to give minimum attenuation may be determined by differentiating α with respect to C and equating to zero.

The result obtained in this case is :—

$$C = \frac{LG}{R}\text{ farads/mile} \qquad (35)$$

In practice, C is normally already greater than the value given by $\frac{LG}{R}$, and to reduce the attenuation it would be necessary to decrease the capacity.

Values of R and G for minimum attenuation.—If either R or G is the only variable, no minimum is found by differentiating and equating to zero. When, however, $R = 0$ and $G = 0$, the attenuation is zero, as can be seen from equation 29; hence R and G should both be kept as small as possible.

Conditions for minimum distortion

If the received signal is not an exact replica of the transmitted signal, the signal is said to be "distorted."

There are three main causes of distortion along a transmission line. Distortion occurs when:—

(1) The characteristic impedance of the line varies with frequency and the line is terminated in an impedance that does not vary with frequency in an identical manner.

(2) The attenuation of the line varies with frequency, so that waves of different frequencies are attenuated by different amounts.

(3) The velocity of propagation varies with frequency so that waves of different frequencies arrive at different times.

Distortion due to Z_0 varying with frequency.—It has been seen that:—

$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \quad \text{(equation 16)}$$

$$Z_0 = \sqrt{\frac{R\left(1 + j\omega \frac{L}{R}\right)}{G\left(1 + j\omega \frac{C}{G}\right)}}$$

It will be seen that, when $LG = CR$, i.e., $\frac{L}{R} = \frac{C}{G}$, then:—

$$\left(1 + j\omega \frac{L}{R}\right) = \left(1 + j\omega \frac{C}{G}\right)$$

$$Z_0 = \sqrt{\frac{R}{G}}\angle 0^\circ = \sqrt{\frac{L}{C}}\angle 0^\circ \qquad (41)$$

In such a case, Z_0 no longer depends on ω, is therefore independent of frequency, and is resistive.

Hence a line for which $\frac{L}{R} = \frac{C}{G}$ can readily be correctly terminated in its characteristic impedance at all frequencies, thus eliminating this form of distortion.

Distortion due to α varying with frequency.—It has been seen that:—

$$\gamma = \sqrt{(R + j\omega L)(G + j\omega C)} \quad \text{(equation 21)}$$

$$\therefore \gamma^2 = RG + j\omega\,(CR + LG) - \omega^2 LC \qquad (42)$$

When $LG = CR$,

$$CR = LG = \sqrt{CRLG}$$

hence $(CR + LG) = 2\sqrt{CRLG}$

From equation 42 —

$$\gamma^2 = RG + 2j\omega\sqrt{RGLC} - \omega^2 LC$$

$$\therefore \gamma^2 = [\sqrt{RG} + j\omega\sqrt{LC}]^2$$

$$\therefore \gamma = \sqrt{RG} + j\omega\sqrt{LC}$$

But $\gamma = \alpha + j\beta$

Hence, when $LG = CR$,

$$\therefore \alpha = \sqrt{RG} \qquad (43)$$

and

$$\beta = \omega\sqrt{LC} \qquad (44)$$

If $\alpha = \sqrt{RG}$, it is independent of frequency, hence there will be no distortion due to the attenuation varying with frequency.

Distortion due to velocity varying with frequency

The velocity of propagation $v = \dfrac{\omega}{\beta}$ (equation 15)

When $\dfrac{L}{R} = \dfrac{C}{G}$

$$\beta = \omega\sqrt{LC}$$

Thus $v = \dfrac{\omega}{\omega\sqrt{LC}} = \dfrac{1}{\sqrt{LC}}$ miles per second (45)

ω has now disappeared from the velocity equation, and hence v* is independent of frequency

The "distortionless condition"

$LG = CR$ is called the *distortionless condition* for a line, for when this relationship holds, the received signal is an exact replica of the sent signal, although reduced in amplitude and delayed by a constant time It will be noted that this condition for minimum distortion is identical with that for minimum attenuation when either L or C may be varied It is evident, then, that the transmission properties of a line can be greatly improved if either L can be increased or C decreased in order that this condition be fulfilled

C depends on the construction of the line or the make-up of the cable and cannot readily be reduced Attempts have therefore been concentrated on efforts to increase L "Loading" is the

---

* When $\beta = \omega\sqrt{LC}$, the group velocity $\dfrac{d\omega}{d\beta}$ also equals $\dfrac{1}{\sqrt{LC}}$, and is therefore independent of frequency

name given to the process whereby the inductance of the line is artificially increased to reduce the attenuation and distortion. There are two main types of loading in general use—continuous loading and lumped loading.

With regard to distortion alone, attempts have been made to satisfy the condition $LG = CR$ by increasing G. These have proved unsatisfactory, however, due to the fact that although increasing G reduces the distortion, it increases the attenuation.

Continuous loading

A tape of iron or some other magnetic material such as mumetal is wound round the conductor to be loaded, thus increasing the

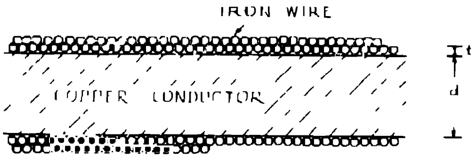


FIG. 739.—Continuous loading.

permeability of the surrounding medium and thereby increasing the inductance (*see* Fig. 739). It may be shown that the increase in inductance is :—

$$L \simeq \frac{\mu}{\frac{d}{n \cdot t} + 1} \text{ mH}$$

where μ = permeability of iron wire.
d = diameter of copper conductor.
n = number of layers.
t = thickness per layer of iron wire.

If $\mu = 200$, $n = 2$, and $t = 0{\cdot}005$ in., then the additional inductance per mile in terms of the diameter d of copper wire is given by Table XXI.

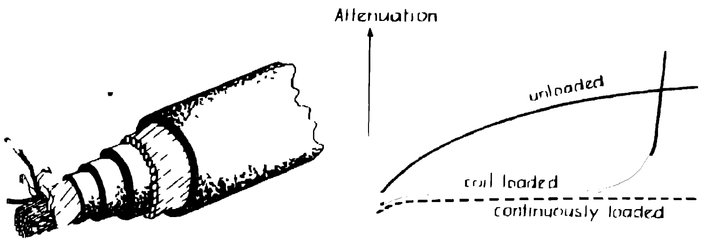
TABLE XXI

Type of conductor	d	Additional inductance per mile
40 lb. copper ..	0·05 in.	34 mH
70 lb. copper ..	0·066 in.	27 mH
100 lb. copper ..	0·079 in.	23·5 mH

This method may be used to give an increase in inductance up to approximately 100 millihenries per mile, but the cost is excessive due to the difficulties in construction. Further disadvantages are the large apparent increase in the primary constant R, due to eddy current losses and hysteresis losses in the magnetic material, and the fact that small differences in mechanical treatment or pressure between the tape and conductor cause large variations in the primary constants.

Continuous loading at the present time is used only on submarine cables, where the problem of making water-tight joints at loading points renders "lumped" loading difficult; further, repair of a break in the cable would probably result in an alteration in the loading coil spacing for that section and hence introduce irregularities.

The continuously loaded cable has the advantage over the lump-loaded cable, that its attenuation increases smoothly with increase in frequency; there is no "cut-off" frequency (*See* Fig. 740*b*.)



(a) CONTINUOUSLY LOADED CABLE

(b) ATTENUATION FREQUENCY CHARACTERISTIC OF CABLE

FIG 740

It has been found unnecessary to use continuous loading over the entire cable to obtain the required reduction in attenuation and distortion. Some submarine cables employ sections of continuously loaded cable separated by sections of unloaded cable, a typical length for the sections being 440 yds. In this way the benefits of continuous loading are obtained but the cost is greatly reduced. This system is known as "patch" loading.

Lumped loading

The inductance of a line may be increased by the introduction of inductance coils at uniform intervals along the line. Provided that the spacing is uniform the line behaves, at all frequencies up to a frequency called the "cut-off" frequency of the line, as if this added inductance were distributed uniformly along it. Above this cut-off frequency, the attenuation increases rapidly. The line, in fact, acts as if it were a low-pass filter.

Provided that a limited frequency range is permissible, this method of loading is more convenient than continuous loading. There is, however, a practical limit to the amount by which the inductance of the line may be increased to reduce attenuation: the inductance (or loading) coils have a certain resistance, and thus increasing L also increases R. Moreover, hysteresis and eddy current losses will occur in loading coils. These will cause a further apparent increase in R, and unless the coil is carefully designed, may introduce distortion. The resistances of several typical loading coils may be seen from Table XXII.

TABLE XXII

Characteristics of typical loading coils

Type	Inductance (mH)	Resistance (Ω)
Pots, loading, 2-coil, No. 2	4·6	2·3
Pots, loading, 2 coil, No. 3	88	10·0
G.P.O. type A88	88	3·0
G.P.O. type B88	88	4·3
G.P.O. type 506	250	5·6
G.P.O. type 582	250	10·5

Fig. 741*a* and *b* show the effect of loading on the attenuation and characteristic impedance of carrier quad cable at various frequencies. They also show the variation of attenuation and characteristic impedance with frequency for several other types of line.

Construction of loading coils

The most important features in loading coil design are low resistance, low core loss, maintenance of circuit balance, avoidance of interference between circuits, and (particularly for field work) small size.

The core is usually toroidal in shape, and made of permalloy or molybdenum-permalloy dust, bound by shellac. This form of core permits the construction of a coil of high inductance, having small dimensions, very low eddy current losses, and negligible external field which might otherwise cause interaction with neighbouring circuits.

The coil is wound of the largest gauge of wire consistent with small size, and each winding is divided into equal parts, so that exactly half the inductance can be inserted into each leg of the circuit. To avoid cross-talk, a high order of accuracy in balancing is necessary, and Fig. 742 shows the method of winding employed to ensure that the two parts of the winding are identical.

Loading coils are usually built into steel "pots," which are made in several standard sizes to accommodate one or more coils.

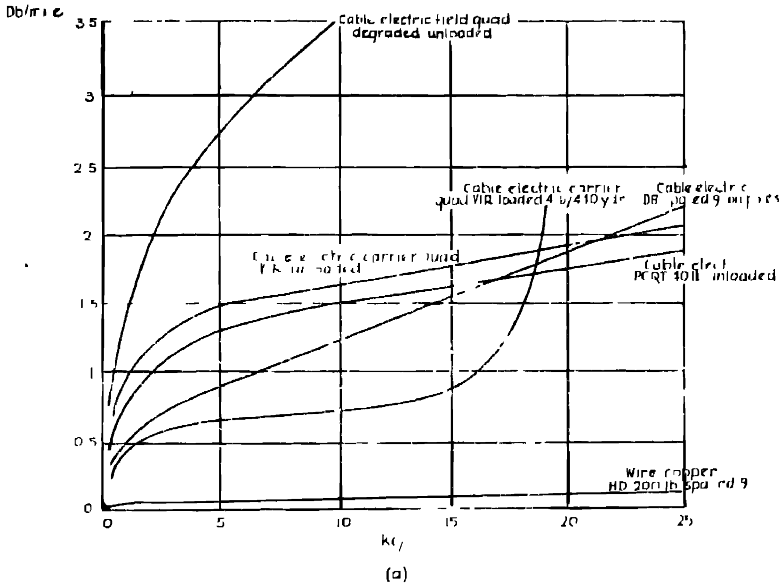


FIG. 741.—(*a*) Variation of attenuation with frequency for some army lines

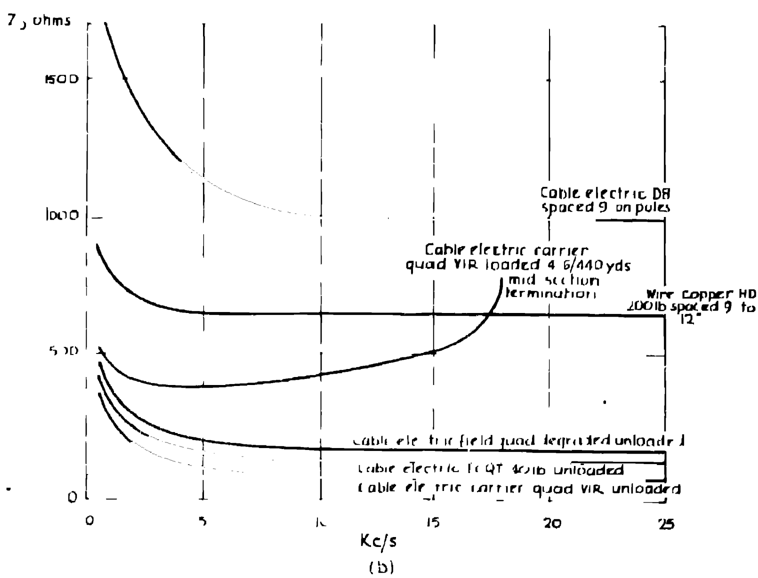


FIG. 741—(*b*) Variation of $|Z_0|$ with frequency for some army lines.

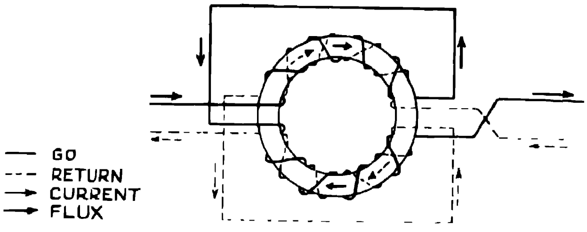


FIG. 742.—Winding of single loading coil, showing directions of current and flux.

In addition to giving the coils protection from the weather and from mechanical damage, the pots also screen the coils from external magnetic fields.

Important considerations in the use of loading coils

When installing loading coils, one must ensure not only that the circuit balance is maintained, but also that the inductances and spacings of the various coils on any circuit are all equal within fairly fine limits. In the case of repeatered circuits, divergencies of more than about 2 per cent. from the mean value of inductance

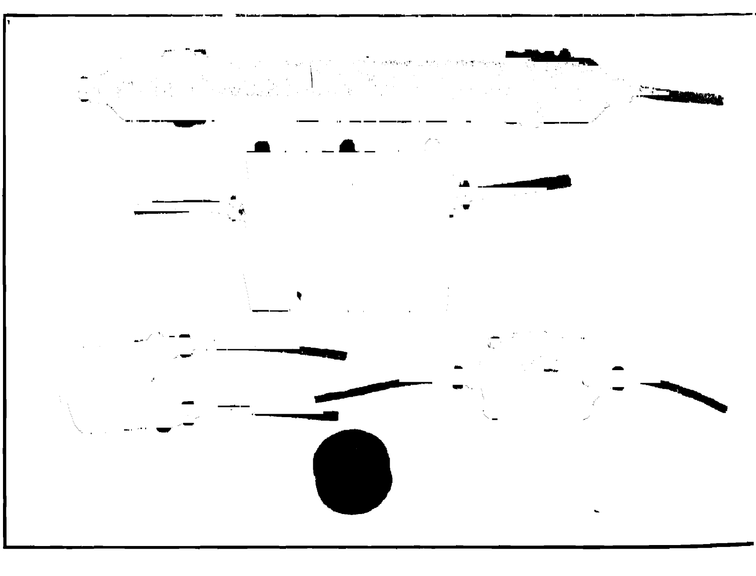


PLATE 36. – Loading coils.

or spacing will render the construction of accurate line balancing networks almost impossible, and so prevent the setting-up of a stable circuit with a good overall T.E., as explained in Chapter 21. When it is impossible to locate a coil at the correct distance, the effective length of a section of line may be artificially increased by a process of " building-out " as described below.

Care must always be taken that no winding is reversed, or it will neutralize the inductance of the other winding of the coil instead of adding to it.

When DC telegraphy is employed over a loaded circuit the telegraph current must not exceed the maximum permissible for the coils in use, or permanent magnetisation may occur, causing a reduction in the effective inductance of the coils and also introducing distortion.

Building-out short loading circuits

When, for geographical and similar reasons, loading pots cannot be located at the correct spacing, short sections can be " built-out " to the correct electrical length. In the case of a long section, an additional loading point must be inserted, and the short section (or sections) resulting can then be built-out.

The principle of building-out is simply the addition of capacity and, if great precision is required, of resistance, at a convenient point in the section. The capacity must be added not only between the two legs of a pair, but also between all wires in a cable, and in

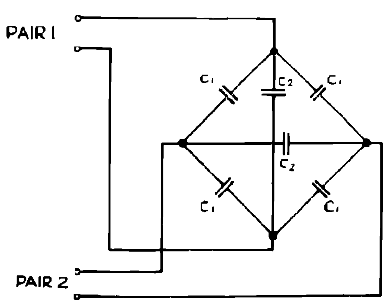


FIG. 743.—Building-out capacity.

the same proportion as the capacities of the cable. Thus, in the case of a quad, six capacities must be added as in Fig. 743.

The capacity C_b to be added is not simply equal to that of the " missing " portion of the section, but is given by the formula :—

$$C_b = C_s - d_1 C \qquad (46)$$

where C_s is the lumped capacity that would simulate the distributed capacity per section,

d_1 is the length of the short section,

and C is the capacity per mile of the line.

In practice, the most convenient method of adding capacity is the use of "stub cables." These are short lengths of cable, usually made with a very high capacity, that can be bridged across or connected in series with the section to be built-out. Series connection adds resistance as well as capacity, but this is not usually necessary; the parallel connection is the more common, as it permits convenient adjustment of the added capacity by varying the length of the "open" end of the stub. Separate resistance can, of course, be added in series with the conductors of the cable if required. When large values of capacity are to be added, several stubs may be connected in parallel to avoid excessive lengths of stub. In either case, the stub must be balanced against cross-talk in the same way as the main cable.

Side circuit and phantom circuit loading

Two metallic pairs between two places *A* and *B* may be utilized to provide three speech circuits by the use of the phantom circuit.

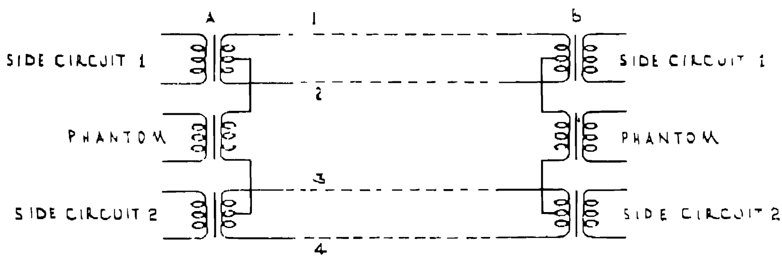


FIG. 744.—Side and phantom circuits.

This is usually tapped off at the line transformers at either end, as in Fig. 744.

Briefly, the reason that speech is possible on the phantom circuit without interference with the side circuits is that current arriving at the centre tap of the line transformer side circuit 1 at *A* splits equally, half travelling to *B* *via* line 1 and half *via* line 2. These two currents will produce no resultant magnetic flux in the iron

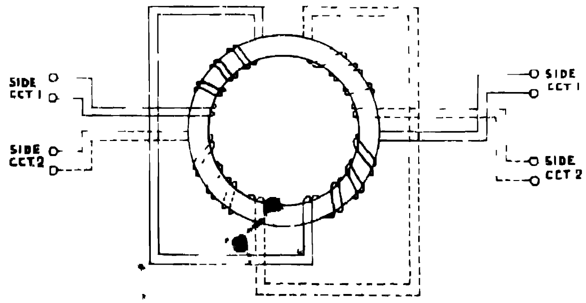


FIG. 745.—Windings of a phantom circuit loading coil.

cores of line transformers at either A or B, and hence no interference with the side circuit 1. The same applies to the return path *via* side circuit 2.

Since the two phantom currents in 1 and 2 are in the same direction, a loading coil such as is shown in Fig. 742 will have zero inductance since the two equal currents will produce equal and opposite magnetic fluxes. This loading coil, although satisfactory for the side circuit, will be useless as far as the phantom circuit is

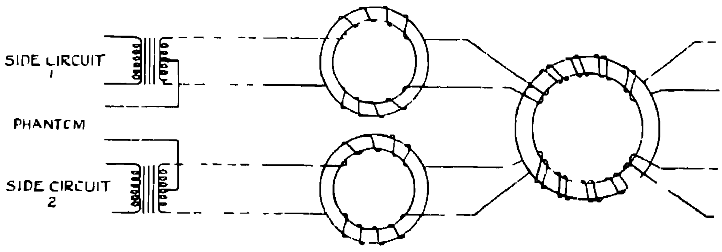


FIG 746.—Arrangement of side and phantom loading coils.

concerned. If it is desired to load the phantom circuit, phantom loading coils must be used (*see* Fig 745). When inserting such a loading coil, side circuit 1 would be treated as one line, and side circuit 2 as the other (*see* Fig. 746).

Cut-off frequency

A lump-loaded line acts as a low-pass filter, since the inductance is lumped instead of being uniformly distributed along the line.

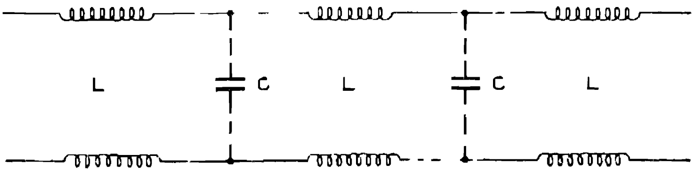


FIG 747.—A loaded line shown as a low-pass filter

The line can be represented approximately by Fig. 747. If the inductance of the line plus loading coil is L_S henries per loading coil section, and the capacity of the line per loading coil section is C_S, then the cut-off frequency f_0 is given by .

$$f_0 = \frac{1}{\pi\sqrt{L_S C_S}} \tag{47}$$

Alternatively, if L is the apparent inductance of the line per mile after loading,

C is the capacity of the line per mile,

and d the loading coil spacing in miles.

Then $L_s = Ld$

$C_s = Cd.$

$$\text{Thus } f_o = \frac{1}{\pi d\sqrt{LC}} \tag{48}$$

Hence the coil spacing d must be less than $\frac{1}{\pi f\sqrt{LC}}$ where f is the highest working frequency.

It will be noted from equation 48 that :—

(1) If the value of the loading coil inductance remains unchanged, the cut-off frequency is inversely proportional to the square root of the loading coil spacing.

(2) If the loading coil spacing remains unchanged, the cut-off frequency is inversely proportional to the square root of the inductance of the loading coil (*see* Fig. 748).

(3) If the loading coil inductance is multiplied by any amount and the spacing is divided by the same amount, there is no change in the cut-off frequency.

(4) If the inductance of each coil, and the spacing between coils are both divided by a factor n, the cut-off frequency is increased by the same factor n.

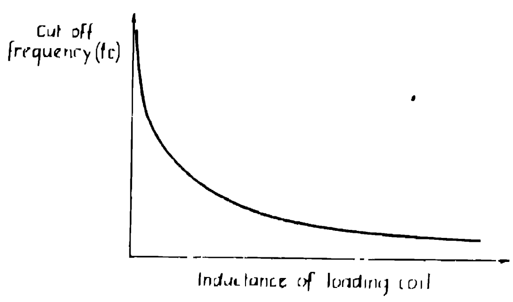


FIG. 748.—Effect on cut-off frequency of varying the inductance of the loading coils with fixed spacing.

Example.—Carrier quad cable, when loaded with 4·6 mH loading coils at 440 yds. spacing, has a cut-off frequency of 24,000 c/s.

(*a*) What will be the cut-off frequency if the spacing is reduced to 110 yds ? *Ans.* 48,000 c/s.

(*b*) What will be the cut-off frequency if 18·4 mH loading coils are used at 440 yds. spacing ? *Ans.* 12,000 c/s.

(*c*) What will be the cut-off frequency if 18·4 mH loading coils are used at 110 yds. spacing ? *Ans.* 24,000 c/s.

(*d*) What will be the cut-off frequency if 2·3 mH loading coils are used at 220 yds. spacing ? *Ans.* 48,000 c/s.

Half-coil and half-section terminations

It has been stated that a loaded line behaves in a similar manner to a low-pass filter. There will therefore be two methods of terminating such a line. If the line is terminated half-way along a loading coil section, an impedance Z_{02} will be obtained corresponding to $Z_{0\pi}$ for a low-pass filter. If the line is terminated in a loading coil having half the normal value, an impedance Z_{01} will be obtained corresponding to Z_{0T} for a low-pass filter (*see* broken curves in Fig. 749).

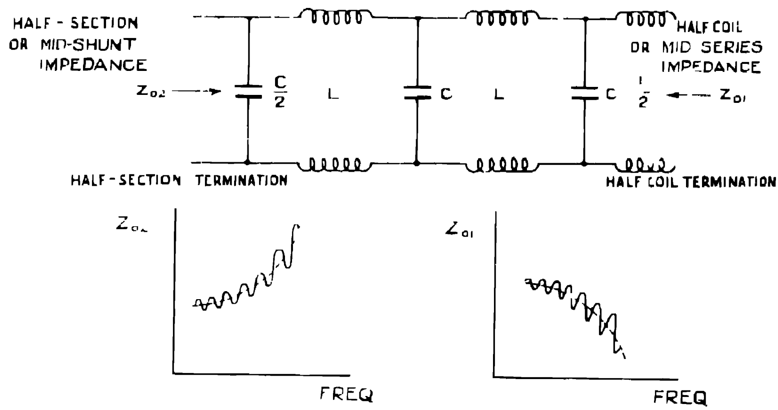


FIG. 749.—Variation in Z_{01} and Z_{02} with frequency.

In practice, the curves obtained for the input impedance of a loaded line, although having the same general shape as those corresponding to a low-pass filter, will not be smooth owing to the presence of points of reflection along the line (*see* full curves in Figs. 749).

Effect of lumped loading on a practical line

It is found in practice that lines can be made almost distortionless by loading only up to a fraction of the theoretical value, the final spacing and size of the coils being a compromise between the attempts to -

(i) obtain a high cut-off frequency,

(ii) use as few loading coils as possible,

(iii) increase the resistance by as small amount as possible.

It has been shown that, to obtain a high cut-off frequency, small loading coils spaced at short intervals must be employed. If loading coil spacing is to be economical, this means that, for a given value of added inductance per mile, a definite limit is set to the cut-off frequency. For this reason loading is most commonly applied to underground cables carrying audio-frequency circuits. A typical form of loading in this case is the insertion of coils of 88 mH at

intervals of 2,000 yds., such a system being designated as "88 mH 2,000 yds." or "88 mH/1·136 miles." It should be noted that the value 88 mH is only a small fraction of that required to satisfy the distortionless condition $L = \frac{CR}{G}$.

Until recently, underground cables on main trunk routes were invariably loaded. In the case of multiple twin cables it was the usual practice to use and to load the phantom circuits. With star-quad cables, the phantoms are seldom used, since the pair-to-pair capacity is too great for satisfactory lumped phantom loading.

With the introduction of carrier systems over underground cables, loading has ceased to be so important. At the higher frequencies used, even an unloaded line approaches the distortionless condition. In any case, all but the lightest of loading is out of the question owing to the high cut-off frequency required. The high attenuation of the unloaded cable is overcome by placing valve amplifiers (repeaters) at frequent intervals along the line. The distortion due to variation of attenuation with frequency is corrected by the use of "attenuation equalisers," and the distortion due to the variation of velocity with frequency is corrected by the use of "phase equalisers." Phase equalisers are used only on the highest grade circuits.

Airline is seldom loaded because the value of L is already approaching that required by the distortionless condition, and the introduction of loading would mean the introduction of further resistance into the circuit.

Loaded underground cables

If an underground cable were loaded up to the value $L = \frac{CR}{G}$, then, as has been shown, Z_0 would be $\sqrt{\frac{R}{G}}$ and α would be $\sqrt{RG}$. In practice such loading is impossible owing to the resistance of the loading coils and to the fact that in order to obtain a reasonable cut-off frequency, the loading coil spacing would be impracticable. Consider the example of an underground cable having the following constants :—

$R = 44$ ohms per mile,
$G = 1$ micromho per mile,
$L = 0 \cdot 001$ henries per mile,
$C = 0 \cdot 065$ microfarads per mile.

In order to attain the distortionless condition, L must be increased to a value L' where :—

$$L' = \frac{0 \cdot 065 \times 10^{-6} \times 44}{10^{-6}} = 2 \cdot 86 \text{ henries per mile.}$$

This is more than 2,800 times the natural inductance of the line.

It is clearly impracticable to introduce such heavy loading; for, supposing that this added inductance could be obtained with negligible added resistance, if it were added in the form of a single loading coil of 2.86 henries at one mile spacing, the cut-off frequency would be much too low for an audio circuit.

In this particular case,

$$f_0 = \frac{1}{\pi d\sqrt{L'C}} = \frac{1}{\pi\sqrt{2.86 \times 0.065 \times 10^{-6}}} = 740 \text{ c/s.}$$

In order to increase the inductance of the line by 2.86 henries per mile and still have a cut-off frequency of, say, 3000 c/s, it would be necessary to load the line using loading coils of 715 mH at intervals of 440 yards. Such heavy loading as this is, however, uneconomical, since a much lighter loading such as 88 mH at intervals of 2 000 yards will give a sufficient reduction both in distortion and in attenuation. In practice the length of the loading section is usually standardised at 2 000 yards and the loading coil inductance is standardised in several values ranging from 250 mH down to 2.3 mH with 88 mH the most commonly used.

If the practical loading of 88 mH at 2 000 yard spacing is applied to the cable under consideration, the resultant inductance per mile is $\left(\frac{88}{1.136} + 1\right) = 78.5$ mH per mile, since the lumped inductance may be regarded as uniformly distributed.

In the range 400 to 3000 c/s

ωL varies between 200 and 1500, while $R = 44$ ohms.

ωC varies between 80×10^{-6} and 1250×10^{-6}, while $G = 1 \times 10^{-6}$ mhos.

Thus $\omega L \gg R$, and $\omega C \gg G$ for a loaded underground cable, i.e., $\frac{R}{\omega L}$ and $\frac{G}{\omega C}$ are both small, very small at high frequencies.

Under these conditions, approximate formulae may be found for Z_0, α and β. These approximations are very useful in practice.

Practical formulae for Z_0 and γ for loaded underground cables

Characteristic impedance Z_0

$$Z_0 = \sqrt[4]{\frac{R^2 + \omega^2 L^2}{G^2 + \omega^2 C^2}} \angle \tfrac{1}{2}\left[\tan^{-1}\frac{G}{\omega C} - \tan^{-1}\frac{R}{\omega L}\right] \text{ (equation 18)}$$

$$\therefore Z_0 = \sqrt[4]{\frac{\omega^2 L^2\left(1 + \frac{R^2}{\omega^2 L^2}\right)}{\omega^2 C^2\left(1 + \frac{G^2}{\omega^2 C^2}\right)}} \angle \tfrac{1}{2}\left[\tan^{-1}\frac{G}{\omega C} - \tan^{-1}\frac{R}{\omega L}\right]$$

$\frac{R}{\omega L}$ and $\frac{G}{\omega C}$ are both very small,

$$\therefore \left(1 + \frac{R^2}{\omega^2 L^2}\right) \simeq 1 \text{ and } \left(1 + \frac{G^2}{\omega^2 C^2}\right) \simeq 1$$

Also $\tan^{-1}\frac{R}{\omega L} \doteq \frac{R}{\omega L}$ radians and $\tan^{-1}\frac{G}{\omega C} \doteq \frac{G}{\omega C}$ radians

$$\therefore Z_0 = \sqrt[4]{\frac{\omega^2 L^2}{\omega^2 C^2}} \angle \tfrac{1}{2}\left(\frac{G}{\omega C} - \frac{R}{\omega L}\right)$$

$$\therefore Z_0 = \sqrt{\frac{L}{C}} \angle \tfrac{1}{2}\left(\frac{G}{\omega C} - \frac{R}{\omega L}\right) \tag{49}$$

The characteristic impedance will be seen to have a modulus that is independent of frequency, and to have only a very small angle.

More approximately, since the angle is very small :—

$$Z_0 \doteq \sqrt{\frac{L}{C}} \angle 0^\circ \tag{50}$$

This last approximation is permissible for heavy loading $\left(\frac{G}{C} - \frac{R}{L} = 0 \text{ is the distortionless condition}\right)$, and also for the higher audio frequencies.

Propagation constant γ

$$\gamma = \sqrt{(R + j\omega L)(G + j\omega C)} \qquad \text{(equation 21)}$$

$$i.e., \ \gamma = \sqrt{j\omega L\left(1 + \frac{R}{j\omega L}\right) j\omega C\left(1 + \frac{G}{j\omega C}\right)}$$

$$= j\omega\sqrt{LC}\left[\left(1 + \frac{R}{j\omega L}\right)\left(1 + \frac{G}{j\omega C}\right)\right]^{\frac{1}{2}}$$

Expanding by the binomial theorem,

$$\gamma = j\omega\sqrt{LC}\left[1 + \frac{R}{2j\omega L} + \ldots\right]\left[1 + \frac{G}{2j\omega C} + \ldots\right]$$

This is permissible because

$$\left|\frac{R}{j\omega L}\right| = \frac{R}{\omega L} \ll 1$$

and $$\left|\frac{G}{j\omega C}\right| = \frac{G}{\omega C} < 1$$

Also, since $\frac{R}{\omega L}$ and $\frac{G}{\omega C}$ are small, second and higher order terms can be neglected, and therefore :—

$$\gamma \doteq j\omega\sqrt{LC}\left[1 + \frac{R}{2j\omega L} + \frac{G}{2j\omega C}\right]$$

$$i.e., \ \gamma \doteq \frac{R}{2}\sqrt{\frac{C}{L}} + \frac{G}{2}\sqrt{\frac{L}{C}} + j\omega\sqrt{LC}$$

The attenuation α of a loaded cable is therefore :—

$$\alpha \doteq \frac{R}{2}\sqrt{\frac{C}{L}} + \frac{G}{2}\sqrt{\frac{L}{C}} \text{ nepers per mile.} \tag{51}$$

Since ω no longer appears in the expression for α, the attenuation will be independent of frequency.

For loaded paper core cables (where G is negligible)

$$\alpha \simeq \frac{R}{2}\sqrt{\frac{C}{L}} = \frac{R}{2|Z_0|} \text{ nepers per mile.} \qquad (52)$$

The phase constant β for a loaded cable is :—

$$\beta \simeq \omega\sqrt{LC} \text{ radians per mile.} \qquad (53)$$

Thus velocity $v^* = \frac{1}{\sqrt{LC}}$, a value that is independent of frequency.

It will thus be seen from equations 51 and 53 that, although the loading is only a small fraction of the value required to make $L = \frac{CR}{G}$, the line approximates to the distortionless condition.

Example.—A 40 lb PCQT underground cable with constants :

$R = 44$ ohms per mile,
$G = 1$ micromho per mile,
$L = 0 \cdot 001$ henries per mile,
$C = 0 \cdot 065$ microfarads per mile,

is loaded with 88 mH loading coils of resistance $3 \cdot 7$ ohms at 2,000 yd. spacing. Find the approximate values of Z_0, α and β.

The total inductance for 2,000 yds. is $(88 + 1 \cdot 136)$ mH., *i.e.*, $89 \cdot 14$ mH.

Therefore the total inductance per mile is :—

$$L' = \frac{89 \cdot 14}{1 \cdot 136} = 78 \cdot 5 \text{ mH.}$$

The total resistance per 2,000 yds. is $(44 \times 1 \cdot 136 + 3 \cdot 7)$ ohms, *i.e.*, $53 \cdot 7$ ohms.

Therefore the total resistance per mile is :—

$$R' = \frac{53 \cdot 7}{1 \cdot 136} = 47 \cdot 3 \text{ ohms.}$$

$$\text{Then } Z_0 \simeq \sqrt{\frac{L'}{C}} = \sqrt{\frac{78 \cdot 5 \times 10^{-3}}{0 \cdot 065 \times 10^{-6}}} = 1100 \text{ ohms (using equation 50)}$$

$$\alpha \simeq \frac{R'}{2}\sqrt{\frac{C}{L'}} + \frac{G}{2}\sqrt{\frac{L'}{C}} = \frac{47 \cdot 3}{2 \times 1100} + 0 \cdot 5 \times 10^{-6} \times 1100$$

i.e., $\alpha \simeq 0 \cdot 022$ nepers per mile $= 0 \cdot 191$ db per mile

and $\beta \simeq \omega\sqrt{L'C} = 10{,}000\sqrt{78 \cdot 5 \times 10^{-3} \times 0 \cdot 065 \times 10^{-6}}$

$= 0 \cdot 715$ at 1600 c/s.

---

* In this case the group velocity $\frac{d\omega}{d\beta}$ is also $\frac{1}{\sqrt{LC}}$ and is independent of frequency.

i.e., $\beta = 0{\cdot}715$ radians per mile at 1600 c/s

and $\beta = 0{\cdot}179$ radians per mile at 400 c/s.

$\therefore v = \dfrac{\omega}{\beta} = 14{,}000$ miles/sec. at both frequencies.

The following table shows these results for partial loading compared with those for the same line unloaded :—

TABLE XXIII

Freq.	Unloaded Z_0	Unloaded α Nepers/mile	Unloaded v miles/sec.	Loaded 88 mH/2,000 yds. Z_0	Loaded 88 mH/2,000 yds. α Nepers/mile	Loaded 88 mH/2,000 yds. v miles/sec
400 c/s	$520\angle -43^\circ$	$0{\cdot}058$	41,000	$1100\angle 0$	$0{\cdot}022$	14,000
1600 c/s	$260\angle -39^\circ$	$0{\cdot}107$	74,000	$1100\angle 0$	$0{\cdot}022$	14,000

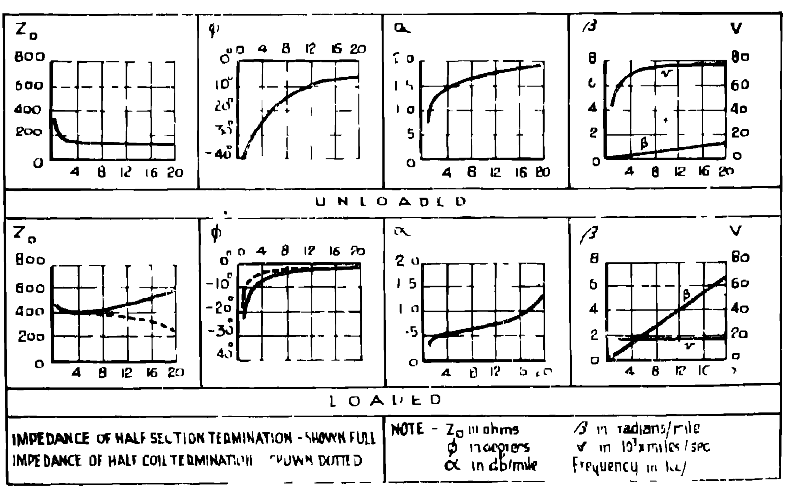


FIG. 750—Secondary constants of carrier quad cable loaded 4·6 mH/440 yds.

Summary of effects of loading

The effects of loading on the secondary constants of a line may be summed up as follows :—

(*a*) The characteristic impedance Z_0 is increased and becomes practically a pure resistance.

(*b*) The attenuation constant α is reduced and becomes practically constant over the working range.

(*c*) The phase constant β is increased, and the velocity of propagation is reduced to a value which is practically constant over the working range.

These effects may be seen from a comparison of the curves given in Fig. 750, which illustrate the secondary constants of carrier quad cable when unloaded and when loaded with 4·6 mH/440 yds.

REFLECTION

So far, current and voltage relationships have only been considered for infinite uniform lines, or uniform lines terminated in their characteristic impedances. If a line, at any point along its length, is joined to some impedance having a value other than Z_0, part of the wave travelling down the line will be reflected back again from the point of discontinuity. In particular, if a line is uniform along its length but is terminated in an impedance Z_R, reflection will occur at the distant end. This reflection will be a maximum when the line is on open circuit ($Z_R = \infty$) or short-circuit ($Z_R = 0$), and will be zero when $Z_R = Z_0$.

Before proceeding further to discuss the current and voltage relationships along such a line, the magnitude of the reflected wave will be considered in more detail in a general case.

Reflection coefficient

In general terms, it may be stated that reflection occurs wherever there is an impedance mis-match between two networks. Consider a generator network *A*, impedance Z_0, working into a load network *B* impedance Z_R (*see* Fig. 751). According to the concept of reflection,

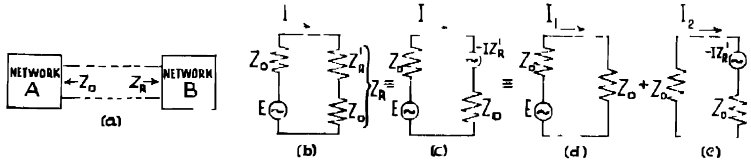


FIG. 751.

an "initial current" I_1 flows from the generator expecting to find a load equal to Z_0. This current in a load Z_R produces a "reflected current" I_2, flowing back from the load to the generator. The resultant current I in the steady state is therefore: —

$$I = I_1 + I_2$$

Applying Thévenin's Theorem, replace network *A* by a generator of EMF *E* and impedance Z_0. Replace network *B* by the two

impedances Z_0 and Z'_R (Fig. 751*b*) such that :–

$$Z_0 + Z'_R = Z_R \tag{54}$$

Applying the Compensation Theorem, replace impedance Z_R by a generator having zero internal impedance and an EMF equal at all times to $-IZ'_R$ (Fig. 751*c*).

Thus $$\frac{E - IZ'_R}{2Z_0} = I$$

$$\therefore \quad E = I(2Z_0 + Z'_R) \tag{55}$$

Applying the Superposition Theorem to Fig. 751*c*, the current may be considered as the sum of two currents: the current I_1 produced by the EMF E, and the current I_2 produced by the EMF $-IZ'_R$ (Figs. 751*d* and *e*).

From Fig. 751*d* :–

$$I_1 = \frac{E}{2Z_0} \tag{56}$$

From Fig. 751*e* :—

$$I_2 = \frac{-IZ'_R}{2Z_0} \tag{57}$$

Dividing (57) by (56) :—

$$\frac{I_2}{I_1} = \frac{-IZ'_R}{E}$$

But from (55) :—

$$E = I(2Z_0 + Z'_R)$$

$$\therefore \quad \frac{I_2}{I_1} = \frac{-IZ'_R}{I(2Z_0 + Z'_R)}$$

$$\therefore \quad \frac{I_2}{I_1} = \frac{-Z'_R}{2Z_0 + Z'_R}$$

$$\therefore \quad \frac{I_2}{I_1} = \frac{Z_0 - Z_R}{Z_0 + Z_R} \text{ (using equation 54)} \tag{58}$$

$\frac{Z_0 - Z_R}{Z_0 + Z_R}$ is called the " reflection coefficient," and gives the ratio of the " reflected " current to the " incident " current. In addition from equation (57),

$$\frac{I_2}{I} = \frac{-Z'_R}{2Z_0}$$

$$= \frac{Z_0 - Z_R}{2Z_0} \tag{59}$$

Therefore $\frac{Z_0 - Z_R}{2Z_0}$ gives the ratio of reflected current to total current flowing.

Derivation of general line equations from reflection considerations

Consider a line of length l, and characteristic impedance Z_0, that is terminated in Z_R at the distant end. Consider a generator of impedance Z_0 and EMF E connected to the sending end (Fig. 752);

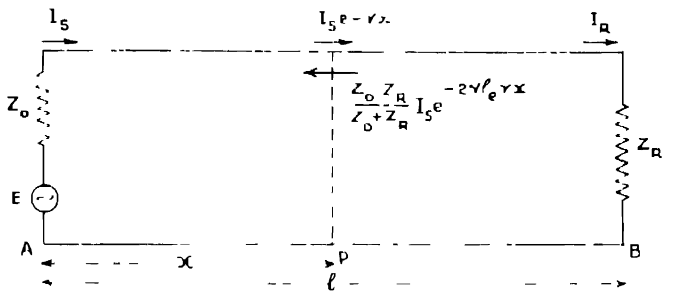


FIG 752 — Reflected current in line terminated in Z_R.

let the sent current be I_S, and the received current be I_R. It is required to find the current and voltage at any point P distant x from the sending end.

The problem cannot be solved by using the infinite line equation directly because the line is neither infinitely long, nor terminated at the receiving end in Z_0.

Consider the current at P.

The generator of EMF E may be considered to send a current wave along the line in the direction A to B, while a return current wave can be considered to be "reflected" back along the line from B to A. The current at P at any instant is the vector sum of these two currents.

Let the total current at P be I.

Let the incident current at P be I_1, and the reflected current at P be I_2.

Let γ be the propagation constant of the line.

Then :—

$$I_1 = I_S e^{-\gamma x}$$
$$= b e^{-\gamma x} \text{ where } b = I_S$$

Received current at B is given by :—

$$I_R = I_S e^{-\gamma l}$$

∴ Reflected current at B is $\frac{Z_0 - Z_R}{Z_0 + Z_R} I_S e^{-\gamma l}$ (*see* equation 58)

$$\therefore \quad I_2 = \frac{Z_0 - Z_R}{Z_0 + Z_R} I_S e^{-\gamma l} e^{-\gamma (l-x)}$$
$$= \frac{Z_0 - Z_R}{Z_0 + Z_R} I_S e^{-2\gamma l} e^{\gamma x}$$

$$i.e. \qquad I_2 = ae^{\gamma x} \text{ where } a = \frac{Z_0 - Z_R}{Z_0 + Z_R} I_S e^{-2\gamma l}$$

The total current I at point P is therefore :—

$$I = I_1 + I_2$$
$$i.e. \qquad I = ae^{\gamma x} + be^{-\gamma x}$$

$be^{-\gamma x}$ represents the wave starting from A and travelling towards B, and $ae^{\gamma x}$ represents the reflected wave starting from B and travelling towards A.

An expression of a similar general form can be derived for the voltage at P, but the mathematical treatment is rather more advanced. This will be considered from a different aspect in Chapter 17.

CHAPTER 17

MATHEMATICAL TREATMENT OF LINE TRANSMISSION

This chapter gives an alternative approach to some aspects of the subject of Line Transmission. In 1893 Kennelly and Steinmetz, working independently, introduced the use of hyperbolic functions of complex numbers in order to simplify the form of the results obtained. This chapter is intended to illustrate the use of these functions. It will be noted that many of the results obtained will have already been derived or assumed in the preceding chapter, but the use of hyperbolic functions will enable a more detailed study to be made.

DERIVATION OF THE GENERAL LINE EQUATIONS

It is required to obtain expressions for current, voltage and impedance, in the steady state, at any point along a line of any length having uniformly distributed electrical constants. Since the line may be terminated in an impedance not equal to its characteristic impedance, the result must take into account the possibility of reflection.

Let the line have length l, characteristic impedance Z_0 and propagation constant γ per mile.

Let R, G, L and C be the primary constants of the line per mile It is assumed that they do not vary with frequency.

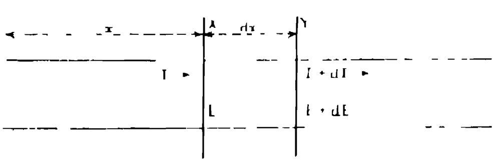


FIG. 753.—Short section XY, distant x from the sending end of a transmission line.

Consider a short section of line XY, of length dx, at a distance x from the sending end (*see* Fig. 753). By making dx very small, the current may be considered constant for voltage calculations, and the voltage constant for current calculations.

At X, let the voltage be E and the current I.

Then at Y, the voltage will be $E + dE$ and the current $I + dI$.

The series impedance of the small section dx will consist of

resistance Rdx and inductance Ldx. The shunt admittance will consist of leakance Gdx and capacitance Cdx.

Since dx is very small, the voltage drop from X to Y may be considered to be due to the current I flowing through the series impedance $Rdx + j\omega Ldx$. The decrease in current from X to Y may be considered to be due to the voltage E being applied to the shunt admittance $Gdx + j\omega Cdx$.

Consider voltage.

Potential difference between X and Y is due to current I flowing through series impedance elements Rdx and $j\omega Ldx$.

Thus $E - (E + dE) = IRdx + Ij\omega Ldx$

$$\therefore \qquad -dE = (R + j\omega L)\, I\, dx$$

$$\therefore \qquad -\frac{dE}{dx} = (R + j\omega L)\, I \tag{1}$$

Consider current.

Current difference between X and Y is due to voltage applied to shunt admittance elements Gdx and $j\omega Cdx$.

Thus $I - (I + dI) = EGdx + Ej\omega Cdx$

$$\therefore \qquad -dI = (G + j\omega C)\, Edx$$

$$\therefore \qquad -\frac{dI}{dx} = (G + j\omega C)\, E \tag{2}$$

To determine the current at distance x from sending end

Differentiate (2) with respect to x

$$-\frac{d^2I}{dx^2} = (G + j\omega C)\frac{dE}{dx}$$

$$\therefore \qquad \frac{d^2I}{dx^2} = (R + j\omega L)(G + j\omega C)\, I \qquad [\text{from } (1)] \tag{3}$$

To simplify notation,

let $\qquad (R + j\omega L)(G + j\omega C) = P$ (say) $\qquad$ (4)

Hence (3) becomes :—

$$\frac{d^2I}{dx^2} = P\,.\,I \tag{5}$$

This is a differential equation the solution of which gives the value of current I at any point distance x down the line, its solution being :—

$$I = ae^{\sqrt{P}x} + be^{-\sqrt{P}x} \tag{6}$$

where a and b are constants.

This may be verified as follows :—

If $\qquad I = ae^{\sqrt{P}x} + be^{-\sqrt{P}x}$

$$\frac{dI}{dx} = a\sqrt{P}e^{\sqrt{P}x} - b\sqrt{P}e^{-\sqrt{P}x}$$

$$\therefore \quad \frac{d^2I}{dx^2} = aPe^{\sqrt{P}x} + bPe^{-\sqrt{P}x}$$

$$i.e., \quad \frac{d^2I}{dx^2} = P\left[ae^{\sqrt{P}x} + be^{-\sqrt{P}x}\right]$$

$$i.e., \quad \frac{d^2I}{dx^2} = PI$$

Before proceeding to a general study of line transmission, some meaning must be given to the constants of equation 6. This can be done by considering a line of infinite length, since the current must become zero as the distance becomes infinite, *i.e.*, $I \to 0$ as $x \to \infty$. Considering equation 6, this means $a = 0$, for the first term *increases* with x. Hence in an infinite line:

$$I = be^{-\sqrt{P}x}$$

If the sending end current is I_S, then I must equal I_S when $x = 0$,

$$\therefore \quad b = I_S$$

Hence the current at any point of an infinite line is:

$$I = I_S e^{-\sqrt{P}x} \tag{7}$$

But the definition of the propagation constant γ of a line is such that, in an infinite line:—

$$I = I_S e^{-\gamma x} \tag{8}$$

Hence, comparing equations 7 and 8, it follows that:—

$$\gamma = \sqrt{P}$$

$$\text{Hence} \quad \gamma = \sqrt{(R + j\omega L)(G + j\omega C)} \tag{9}$$

Thus γ may be determined from the values of the primary line constants. This verifies the result obtained by another method in the last chapter.

To determine the voltage at distance x from sending end (2)

The voltage E can be obtained from equation 2.

$$E = -\frac{1}{G + j\omega C} \cdot \frac{dI}{dx}$$

$$= -\frac{1}{G + j\omega C} \cdot I_S(-\gamma)\, e^{-\gamma x}$$

$$= I_S \cdot \frac{\gamma}{G + j\omega C} \cdot e^{-\gamma x}$$

But $\gamma = \sqrt{(R + j\omega L)(G + j\omega C)}$ from equation 9.

$$\therefore \quad E = I_S \sqrt{\frac{R + j\omega L}{G + j\omega C}}\, e^{-\gamma x} \tag{10}$$

The sending-end voltage E_S can be obtained by putting $x = 0$.

$$\text{hence} \quad E_S = I_S \sqrt{\frac{R + j\omega L}{G + j\omega C}}$$

i.e.,
$$\frac{E_s}{I_s} = \sqrt{\frac{R + j\omega L}{G + j\omega C}}$$

But by definition the ratio $\frac{E_s}{I_s}$ in an infinite line is its characteristic impedance Z_0.

Hence
$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \tag{11}$$

This again verifies a previous result.

These results were obtained by considering an infinite line; the general case of a finite line is given by equation 6, which may now be written :—

$$I = ae^{\gamma x} + be^{-\gamma x} \tag{12}$$

Equation 12 can now be expressed using hyperbolic functions.

Put $e^{\gamma x} = \cosh \gamma x + \sinh \gamma x$

and $e^{-\gamma x} = \cosh \gamma x - \sinh \gamma x$.

Equation 12 therefore now becomes :—

$$I = (a + b) \cosh \gamma x + (a - b) \sinh \gamma x.$$

This may be further simplified by putting $a + b = A$, and $a - b = B$, giving

$$I = A \cosh \gamma x + B \sinh \gamma x. \tag{13}$$

The voltage may now be found in these terms from equation 2.

$$E = -\frac{1}{G + j\omega C} \cdot \frac{dI}{dx}$$

$$= -\frac{1}{G + j\omega C} (A\gamma \sinh \gamma x + B\gamma \cosh \gamma x)$$

$$= -\frac{\gamma}{G + j\omega C} (A \sinh \gamma x + B \cosh \gamma x)$$

$$E = -Z_0 (A \sinh \gamma x + B \cosh \gamma x) \tag{14}$$

since $\gamma = \sqrt{(R + j\omega L)(G + j\omega C)}$ and $Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}}$

Equations 13 and 14 may be used to find the current and voltage at any point along a line, if the constants A and B can be determined. This is usually done from a knowledge of the conditions at one end of the line.

Determination of constants *A* and *B*

(a) *If conditions at sending end of the line are known.*

Let I_s be the current at the sending end, and E_s be the voltage at the sending end.

But at sending end $x = 0$.

Thus equation 13 becomes :

$$I_s = A \cosh \gamma 0 + B \sinh \gamma 0$$

i.e., $I_S = A \times 1 + B \times 0$

i.e., $A = I_S$ (15)

Similarly, equation 14 becomes :—

$$E_S = -BZ_0 \cosh \gamma 0 - AZ_0 \sinh \gamma 0$$

i.e., $E_S = -BZ_0 \times 1 - AZ_0 \times 0$

i.e., $E_S = -BZ_0$

i.e., $B = -\dfrac{E_S}{Z_0}$ (16)

Thus A and B are expressible in terms of the current and voltage at the sending end. Substituting these values for A and B in equations 13 and 14 gives :—

$$I = I_S \cosh \gamma x - \frac{E_S}{Z_0} \sinh \gamma x \qquad (17)$$

and $E = E_S \cosh \gamma x - I_S Z_0 \sinh \gamma x$ (18)

(b) *If conditions at receiving end of the line are known.*

Let I_R be the current at the receiving end, and E_R the voltage at the receiving end.

But at the receiving end $x = l$.

Thus equation 13 becomes :

$$I_R = A \cosh \gamma l + B \sinh \gamma l. \qquad (19)$$

Similarly, equation 14 becomes :—

$$E_R = -BZ_0 \cosh \gamma l - AZ_0 \sinh \gamma l \qquad (20)$$

From equations 19 and 20 :

$$\frac{I_R - A \cosh \gamma l}{E_R + AZ_0 \sinh \gamma l} = \frac{B \sinh \gamma l}{-BZ_0 \cosh \gamma l}$$

$$\therefore\ -I_R Z_0 \cosh \gamma l + AZ_0 \cosh^2 \gamma l = E_R \sinh \gamma l + AZ_0 \sinh^2 \gamma l$$

$$\therefore\quad A = \frac{E_R}{Z_0} \sinh \gamma l + I_R \cosh \gamma l \qquad (21)$$

Similarly :—

$$\frac{I_R - B \sinh \gamma l}{E_R + BZ_0 \cosh \gamma l} = \frac{A \cosh \gamma l}{-AZ_0 \sinh \gamma l}$$

$$\therefore\ -I_R Z_0 \sinh \gamma l + BZ_0 \sinh^2 \gamma l = E_R \cosh \gamma l + BZ_0 \cosh^2 \gamma l$$

$$\therefore\quad B = -I_R \sinh \gamma l - \frac{E_R}{Z_0} \cosh \gamma l \qquad (22)$$

Substitute in equation 13 for A and B :—

$$I = \left[I_R \cosh \gamma l + \frac{E_R}{Z_0} \sinh \gamma l \right] \cosh \gamma x - \left[I_R \sinh \gamma l + \frac{E_R}{Z_0} \cosh \gamma l \right] \sinh \gamma x$$

i.e.,
$$\boxed{I = I_R \cosh \gamma (l - x) + \frac{E_R}{Z_0} \sinh \gamma (l - x)} \tag{23}$$

Substitute in equation 14 for A and B :—

$$E = \left[I_R Z_0 \sinh \gamma l + E_R \cosh \gamma l \right] \cosh \gamma x$$
$$- \left[I_R Z_0 \cosh \gamma l + E_R \sinh \gamma l \right] \sinh \gamma x$$

i.e.,
$$\boxed{E = E_R \cosh \gamma (l - x) + I_R Z_0 \sinh \gamma (l - x)} \tag{24}$$

Equations 17 and 18 are the general line equations expressing respectively the current and voltage at a point distant x from the sending end in terms of the sent current and the sent voltage.

Equations 23 and 24 are the general line equations in another form, but this time the current and voltage at a point distant x from the sending end are expressed in terms of the received current and voltage. Clearly then, equations 23 and 24 can be applied only to a line of finite length, whilst equations 17 and 18 apply also to an infinite line.

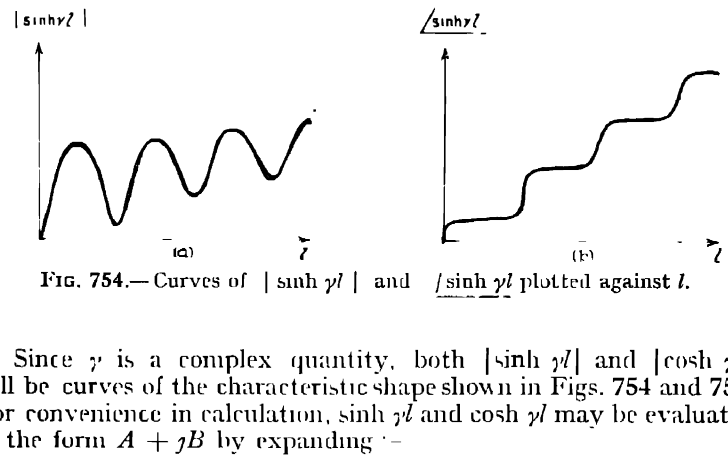


FIG. 754.—Curves of $|\sinh \gamma l|$ and $/\sinh \gamma l$ plotted against l.

Since γ is a complex quantity, both $|\sinh \gamma l|$ and $|\cosh \gamma l|$ will be curves of the characteristic shape shown in Figs. 754 and 755. For convenience in calculation, $\sinh \gamma l$ and $\cosh \gamma l$ may be evaluated in the form $A + jB$ by expanding :—

$$\sinh \gamma l = \sinh (\alpha + j\beta) l$$
$$= \sinh \alpha l \,.\, \cosh j\beta l + \cosh \alpha l \,.\, \sinh j\beta l$$

i.e.,
$$\sinh \gamma l = \sinh \alpha l \,.\, \cos \beta l + j \,.\, \cosh \alpha l \,.\, \sin \beta l \tag{25}$$

and
$$\cosh \gamma l = \cosh (\alpha + j\beta) l$$
$$= \cosh \alpha l \,.\, \cosh j\beta l + \sinh \alpha l \,.\, \sinh j\beta l$$

i.e.,
$$\cosh \gamma l = \cosh \alpha l \,.\, \cos \beta l + j \,.\, \sinh \alpha l \,.\, \sin \beta l \tag{26}$$

Example.—

$$\sinh (0{\cdot}6 + j \,.\, 2{\cdot}9) = \sinh 0{\cdot}6 \,.\, \cos 2{\cdot}9 + j \,.\, \cosh 0{\cdot}6 \,.\, \sin 2{\cdot}9$$
$$= - 0{\cdot}6367 \,.\, 0{\cdot}9710 + j \,.\, 1{\cdot}1855 \,.\, 0{\cdot}2392$$
$$= - 0{\cdot}6183 + j \,.\, 0{\cdot}2836 \quad \textit{Ans.}$$

$$\cosh (0\cdot6 + j\,.\,2\cdot9) = \cosh 0\cdot6\,.\,\cos 2\cdot9 + j\,.\,\sinh 0\cdot6\,.\,\sin 2\cdot9$$
$$-1\cdot1855\,.\,0\cdot9710 + j\,.\,0\cdot6367\,.\,0\cdot2392$$
$$= -1\cdot152 + j\,.\,0\cdot1524 \quad \textit{Ans.}$$

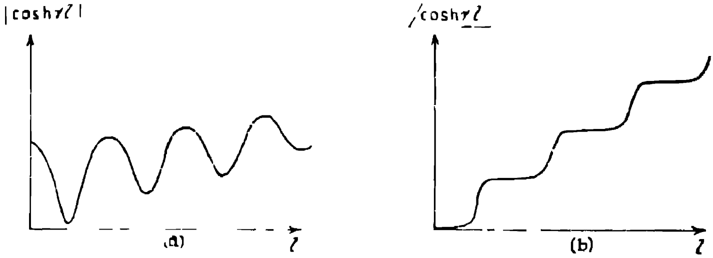


FIG. 755.—Curves of $|\cosh \gamma l|$ and $\angle\cosh \gamma l$ plotted against l

Example

An open-wire line has at 1000 c/s a characteristic impedance Z_0 of $730\,\angle -11°$ and a propagation constant $\gamma = 0\cdot012 + j\,0\cdot058$.

When 2 volts are applied to the sending end, a current of 4 milliamps flows. What will be the current at the distant end 50 miles away?

It will be noted that the value of the terminating impedance is not given. If required it could be calculated from the information given. From equation 17

$$I = I_S \cosh \gamma x - \frac{E_S}{Z_0} \sinh \gamma x$$

$$\therefore \quad I_R = \frac{4}{1000} \cosh 50\,(0\cdot012 + j\,0\cdot058)$$
$$- \frac{2}{730\angle -11} \sinh 50\,(0\cdot012 + j\,0\cdot058)$$
$$= 0\cdot004\,.\,\cosh (0\cdot6 + j\,2\cdot9)$$
$$-(0\cdot002689 + j\,.\,0\cdot0005229) \sinh (0\cdot6 + j\,.\,2\cdot9)$$
$$= 0\cdot004\,(-1\cdot152 + j\,.\,0\cdot1524)$$
$$-(0\cdot002689 + j\,0\cdot0005229)\,.\,(-0\cdot6183 + j\,.\,0\cdot2836)$$
$$= -0\cdot002797 + j\,0\cdot00017$$
$$= 0\cdot0028\,\angle 176°\,35'$$

Thus the current at the receiving end of the line is 2·8 milliamps, leading by 176° 35′ on the sent current. ***Ans.***

APPLICATION OF THE GENERAL LINE EQUATIONS TO PARTICULAR CASES

Case (i). Finite line of length l, terminated in Z_0

At the distant end (i.e., where $x = l$), let the current be I_R and the voltage E_R.

Now E_R can be found in terms of the sent current and voltage by substituting $x = l$ in equation 18

$$\text{i.e.,} \quad E_R = E_S \cosh \gamma l - I_S Z_0 \sinh \gamma l \tag{27}$$

and I_R is found by putting $x = l$ in (17)

$$\text{i.e.,} \quad I_R = I_S \cosh \gamma l - \frac{E_S}{Z_0} \sinh \gamma l \tag{28}$$

$$\text{Thus} \quad \frac{E_R}{I_R} = Z_0 \frac{E_S \cosh \gamma l - I_S Z_0 \sinh \gamma l}{I_S Z_0 \cosh \gamma l - E_S \sinh \gamma l}$$

But since the terminating impedance is Z_0 it follows that :

$$\frac{E_R}{I_R} = Z_0$$

$$\text{Thus} \quad Z_0 \frac{E_S \cosh \gamma l - I_S Z_0 \sinh \gamma l}{I_S Z_0 \cosh \gamma l - E_S \sinh \gamma l} = Z_0$$

$$\therefore \quad E_S \cosh \gamma l - I_S Z_0 \sinh \gamma l = I_S Z_0 \cosh \gamma l - E_S \sinh \gamma l$$

$$\therefore \quad \frac{E_S}{I_S} = Z_0 \frac{\cosh \gamma l + \sinh \gamma l}{\cosh \gamma l + \sinh \gamma l}$$

$$\therefore \quad \frac{E_S}{I_S} = Z_0 \tag{29}$$

But $\frac{E_S}{I_S}$ is the input impedance of the line

$$\text{Thus} \quad Z_{IN} = Z_0 \tag{30}$$

That is to say, the input impedance of a finite line terminated in its characteristic impedance Z_0 is the characteristic impedance Z_0.

Now consider the general equation for current I.

From equation 17 :—

$$I = I_S \cosh \gamma x - \frac{E_S}{Z_0} \sinh \gamma x$$

$$\text{But} \quad Z_0 = \frac{E_S}{I_S} \quad \text{(from equation 29)}$$

Substituting for Z_0 :—

$$\begin{aligned} I &= I_S \cosh \gamma x - E_S \cdot \frac{I_S}{E_S} \sinh \gamma x \\ &= I_S (\cosh \gamma x - \sinh \gamma x) \\ &= I_S \cdot e^{-\gamma x} \end{aligned} \tag{31}$$

This is the same expression as that giving the current at a point distant x along an infinite line.

Similarly, the general equation for the voltage E can be found:—

$$E = E_S \cosh \gamma x - I_S Z_0 \sinh \gamma x.$$

But $\quad Z_0 = \frac{E_S}{I_S}$ (from equation 29)

$$\therefore \quad E = E_S \cosh \gamma x - I_S \cdot \frac{E_S}{I_S} \sinh \gamma x$$

$$= E_S (\cosh \gamma x - \sinh \gamma x)$$

$$\therefore \quad E = E_S \, e^{-\gamma x} \tag{32}$$

which is the same expression as that giving the voltage at a point distant x along an infinite line.

Hence with regard to current, voltage and impedance, a short line terminated in its characteristic impedance behaves as an infinite line. This verifies the assumption made in the previous chapter.

Case (ii). Finite line of length l, open-circuited at distant end

In the case of an open circuited line, the current at the distant end (I_R) is zero.

Thus when $x = l$, equation 17 becomes :—

$$0 = I_S \cosh \gamma l - \frac{E_S}{Z_0} \sinh \gamma l$$

$$\therefore \quad \frac{E_S}{I_S} = Z_0 \frac{\cosh \gamma l}{\sinh \gamma l} = Z_0 \coth \gamma l$$

But $\frac{E_S}{I_S}$ is the input impedance of the line. In the special case of an open-circuited line call it Z_{OC}.

Then $Z_{OC} = Z_0 \coth \gamma l. \quad (33)$

Note that when γl is very large, $\coth \gamma l \to 1$. Hence Z_{OC} approaches Z_0 if l is made very large

Case (iii). Finite line of length l, short-circuited at the distant end

In the case of a short-circuited line, the voltage at the distant end (E_R) is zero.

Thus when $x = l$, equation 27 becomes :—

$$0 = E_S \cosh \gamma l - I_S Z_0 \sinh \gamma l$$

$$i.e., \quad \frac{E_S}{I_S} = Z_0 \frac{\sinh \gamma l}{\cosh \gamma l} = Z_0 \tanh \gamma l$$

But $\frac{E_S}{I_S}$ is the input impedance of the line. In the special case of a short-circuited line, call it Z_{SC}.

Then $Z_{SC} = Z_0 \tanh \gamma l. \quad (34)$

Note that when γl is very large, $\tanh \gamma l \to 1$. Hence Z_{SC} approaches Z_0 if l is made very large.

Multiply equations 33 and 34 :—

$$Z_{OC} \cdot Z_{SC} = Z_0 \coth \gamma l \cdot Z_0 \tanh \gamma l.$$

$$\therefore Z_{OC} \cdot Z_{SC} = Z_0^2$$

Hence $$Z_0 = \sqrt{Z_{OC} \cdot Z_{SC}} \qquad (35)$$

Thus for any uniform and symmetrical line the characteristic impedance is the geometric mean of the open- and short-circuit impedances. This verifies the result obtained in the previous chapter.

Dividing equation 34 by equation 33 gives :—

$$\frac{Z_{SC}}{Z_{OC}} = \frac{Z_0 \tanh \gamma l}{Z_0 \coth \gamma l} = \tanh^2 \gamma l$$

$$\tanh \gamma l = \sqrt{\frac{Z_{SC}}{Z_{OC}}} \qquad (36)$$

Example.—

A line 10 miles long has the following constants :—

$Z_0 = 600 \angle 0°$

$\alpha = 0 \cdot 1$ nepers per mile.

$\beta = 0 \cdot 05$ radians per mile.

Find the received current and voltage when 20 milliamps are sent into one end, and the receiving end is short-circuited.

The general line equation for voltage (equation 18) states :

$$E = E_S \cdot \cosh \gamma x - I_S Z_0 \sinh \gamma x$$

But $E = 0$ at $x = 10$ miles, due to the short-circuit.

$$\therefore \quad 0 = E_S \cdot \cosh \gamma \cdot 10 - I_S Z_0 \cdot \sinh \gamma \cdot 10$$

$$\therefore \quad E_S = \frac{I_S Z_0 \cdot \sinh \gamma \cdot 10}{\cosh \gamma \cdot 10}$$

The general line equation for current (equation 17) states : —

$$I = I_S \cdot \cosh \gamma \cdot x - \frac{E_S}{Z_0} \cdot \sinh \gamma \cdot X$$

$$= 20 \cdot \cosh \gamma \cdot 10 - \frac{I_S Z_0 \cdot \sinh \gamma \cdot 10}{Z_0 \cdot \cosh \gamma \cdot 10} \cdot \sinh \gamma \cdot 10$$

$$= \frac{20}{\cosh \gamma \cdot 10} (\cosh^2 \gamma \cdot 10 - \sinh^2 \gamma \cdot 10)$$

$$= \frac{20}{\cosh (1 + j \cdot 0 \cdot 5)}$$

$$= \frac{20}{1 \cdot 354 + j \cdot 0 \cdot 564}$$

$$= 12 \cdot 6 - j \cdot 5 \cdot 25$$

$$= 13 \cdot 65 \angle -22° \, 30' \quad \textit{Ans.}$$

Thus the received voltage is zero, and the received current is 13·65 milliamps, lagging behind the sent current by 22° 30′.

Practical application of the above formulae

Equations 35 and 36 enable the primary and secondary line constants of a transmission line to be calculated from the measured

values of the open- and short-circuit impedances of a known length of line.

For suppose the modulus and angle of both Z_{oc} and Z_{sc} are found by measurement, then Z_0 will be determined from equation 35, which is a vector equation, and will give both the modulus and angle of Z_0. Similarly, the right-hand side of equation 36 will be a vector quantity, which may be changed into the rectangular form. *i.e.*, $\tanh \gamma l = A + jB$ where A and B are known. (37)

It has been shown (Chapter 2, page 97) that in this case :—

$$\tanh 2\alpha l = \frac{2A}{1 + A^2 + B^2} \tag{38}$$

$$\text{and } \tan 2\beta l = \frac{2B}{1 - (A^2 + B^2)} \tag{39}$$

From equations 38 and 39 (*see also* Appendix II), αl and βl may be deduced, and if the length of the line is known, α and β may be found. αl may be determined from tables, or as follows :—

$$\text{Let } \tanh 2\alpha l = \frac{2A}{1 + A^2 + B^2} = C$$

$$\text{Then} \qquad \frac{e^{4\alpha l} - 1}{e^{4\alpha l} + 1} = C$$

$$\therefore \qquad e^{4\alpha l}(1 - C) = 1 + C$$

$$\therefore \qquad e^{4\alpha l} = \frac{1 + C}{1 - C}$$

$$\therefore \qquad \alpha l = \frac{1}{4}\log_e\left(\frac{1 + C}{1 - C}\right) \tag{40}$$

From β, the wavelength λ $\left(= \frac{2\pi}{\beta}\right)$ and the velocity of propagation v $\left(= \frac{\omega}{\beta}\right)$ may be calculated.

Further, γ is now known ($\gamma = \alpha + j\beta$), and multiplying equations 9 and 11 :—

$$Z_0 \gamma = R + j\omega L \tag{41}$$

Also, by dividing 9 by 11 :—

$$\frac{\gamma}{Z_0} = G + j\omega C \tag{42}$$

From 41 and 42 the four primary constants may be determined.

3*Example.*—

Impedance measurements made on a 440-yard length of field quad cable at 1600 c/s ($\omega = 10{,}000$) under open-circuit and short-circuit conditions gave the following results :—

$$Z_{oc} = 2460 \,\underline{/-86^\circ 30'}\,; \quad Z_{sc} = 21 \cdot 5 \,\underline{/14^\circ}$$

Calculate Z_0, α, β, R, G, L and C.

From (35) : $Z_0 = \sqrt{Z_{oo} \cdot Z_{so}} = \sqrt{21 \cdot 5 \times 2460} \angle -36^\circ\ 15'$

i.e., $Z_0 = 230 \angle -36^\circ\ 15'$

From (36) : $\tanh \gamma l = \sqrt{\dfrac{Z_{so}}{Z_{oo}}} = \sqrt{\dfrac{21 \cdot 5}{2460}} \angle 50^\circ\ 15' = 0 \cdot 0935 \angle 50^\circ\ 15'$

$$= 0 \cdot 0598 + j0 \cdot 0719 = A + jB$$

From (38) : $\tanh 2\alpha l = \dfrac{2A}{1 + A^2 + B^2} = \dfrac{0 \cdot 1196}{1 \cdot 009} = 0 \cdot 1185$

$\therefore$ From (40) : $\alpha l = \dfrac{1}{4} \log_e \dfrac{1 + 0 \cdot 1185}{1 - 0 \cdot 1185} = \dfrac{1}{4} \log_e \dfrac{1 \cdot 1185}{0 \cdot 8815} = 0 \cdot 060$

But $l = 0 \cdot 25$ miles

$\therefore$ $\alpha = 0 \cdot 240$ nepers/mile

From (39) : $\tan 2\beta l = \dfrac{2B}{1 - (A^2 + B^2)} = \dfrac{0 \cdot 1438}{0 \cdot 9912} = 0 \cdot 1452$

$\therefore$ $2\beta l = 8^\circ\ 16' \pm n\pi$, where n is zero or a positive integer.

i.e., $2\beta l = 0 \cdot 1443 \pm n\pi$ radians.

But $l = 0 \cdot 25$ miles

$\therefore$ $\beta = 0 \cdot 289 \pm 2n\pi$ radians/mile.

In order to determine the correct value of n, it is necessary to make one more observation.

(i) *Either* estimate the approximate velocity of propagation in the cable at the frequency considered,

(ii) *or* estimate the loop resistance per mile of the line.

These two methods of approach will be taken in turn.

(i) The velocity of propagation for airline at 1600 c/s is approximately 170,000 miles per second, and for loaded cable approximately 10,000 miles per second. For field quad cable it will be somewhere between these values estimate it at roughly 50,000 miles per second.

Since $v = \dfrac{\omega}{\beta}$, $v \doteqdot 50{,}000$, and $\omega = 10{,}000$,

$\therefore$ $\beta = 0 \cdot 2$ radians per mile (approx.)

$\therefore$ $n = 0$, and the correct value of β is :—

$\beta = 0 \cdot 289$ radians per mile.

(ii) Measuring the DC resistance of the 440 yard loop, the result was $20 \cdot 4$ ohms, giving a DC resistance of $81 \cdot 6$ ohms per mile loop.

The method is to calculate the AC loop resistance R using $\beta = 0 \cdot 289 \pm 2n\pi$ and see which value of n gives the nearest agreement between R and the measured DC loop resistance.

$$\text{Thus} \quad \gamma = 0{\cdot}240 + j\,(0{\cdot}289 \pm 2n\pi)$$
$$Z_0 = 230 \angle{-36^\circ\,15'} = 186 - j\,136$$
$$R + j\omega L = \gamma Z_0 = [186 - j\,136]\,[0{\cdot}240 + j\,(0{\cdot}289 \pm 2n\pi)]$$
$$\therefore \quad R = 186\,.\,0{\cdot}240 + 136\,(0{\cdot}289 \pm 2n\pi)$$
$$= 44{\cdot}6 + 39{\cdot}3 \pm 854n$$
$$= 83{\cdot}9 \pm 854n$$

Since the DC loop resistance is 81·6 ohms, the correct value of n is $n = 0$, giving $\beta = 0{\cdot}289$ as before.

Having established the correct value of β, proceed as follows :—

$$\gamma = 0{\cdot}240 + j\,0{\cdot}289 \quad 0{\cdot}3756 \angle{50^\circ\,26'}$$
$$Z_0 = 230 \angle{-36^\circ\,15'}$$
$$\therefore R + j\omega L = \gamma Z_0 = 230 \times 0{\cdot}3756 \angle{14^\circ\,11'}$$
$$= 83{\cdot}8 + j\,21{\cdot}0$$

Hence $\quad R = 83{\cdot}8$ ohms per mile

and $\quad \omega L = 21{\cdot}0$

$\therefore \quad L = 2{\cdot}1$ mH per mile.

Similarly . –

$$G + j\omega C = \frac{\gamma}{Z_0} = \frac{0{\cdot}3756}{230} \angle{86^\circ\,41'}$$
$$= 0{\cdot}000094 + j\,0{\cdot}00163$$

$\therefore \quad G = 0{\cdot}000094$ mhos per mile.

and $\quad \omega C = 0{\cdot}00163$

$\therefore \quad C = 0{\cdot}163\ \mu$F per mile.

Thus, for the cable considered :—

$$\left.\begin{aligned} R &= 83{\cdot}8 \text{ ohms per mile} \\ L &= 2{\cdot}1 \text{ mH per mile} \\ G &= 94\ \mu\text{mhos per mile} \\ C &= 0{\cdot}163\ \mu\text{F per mile.} \end{aligned}\right\} \quad \textit{Ans.}$$

GENERAL CASE OF A FINITE LINE TERMINATED IN AN IMPEDANCE Z_R

Consider the case of a finite line of length l, terminated in any impedance Z_R. Let the received current and voltage be I_R and E_R respectively.

Then, from equation 17, putting $x = l$:—

$$I_R = I_s \cosh \gamma l - \frac{E_s}{Z_0} \sinh \gamma l$$

and, from equation 18, when $x = l$:—

$$E_R = E_S \cosh \gamma l - I_S Z_0 \sinh \gamma l$$

But $\dfrac{E_R}{I_R} = Z_R$ since the line is terminated at the distant end in Z_R.

Thus
$$\frac{E_S \cosh \gamma l - I_S Z_0 \sinh \gamma l}{I_S \cosh \gamma l - \dfrac{E_S}{Z_0} \sinh \gamma l} = Z_R$$

$$\therefore \quad I_S Z_R \cosh \gamma l - \frac{E_S Z_R}{Z_0} \sinh \gamma l = E_S \cosh \gamma l - I_S Z_0 \sinh \gamma l$$

$$\therefore I_S Z_0 Z_R \cosh \gamma l - E_S Z_R \sinh \gamma l = E_S Z_0 \cosh \gamma l - I_S Z_0^2 \sinh \gamma l$$

The input impedance Z_{IN} is given by :—

$$Z_{IN} = \frac{E_S}{I_S} = Z_0 \cdot \frac{Z_R \cosh \gamma l + Z_0 \sinh \gamma l}{Z_0 \cosh \gamma l + Z_R \sinh \gamma l} \tag{43}$$

Alternatively :—

$$Z_{IN} = Z_0 \cdot \frac{Z_R + Z_0 \tanh \gamma l}{Z_0 + Z_R \tanh \gamma l} \tag{44}$$

Variation of input impedance with frequency

Equation 43 may be written :—

$$Z_{IN} = Z_0 \cdot \frac{\cosh \gamma l + \dfrac{Z_0}{Z_R} \sinh \gamma l}{\dfrac{Z_0}{Z_R} \cosh \gamma l + \sinh \gamma l}$$

Put $\quad \dfrac{Z_0}{Z_R} = \tanh \theta = \tanh (p + jq)$

This is permissible, since values of p and q can be found to give any value for $\tanh \theta$.

Then
$$\begin{aligned} Z_{IN} &= Z_0 \cdot \frac{\cosh \gamma l + \tanh \theta \sinh \gamma l}{\tanh \theta \cosh \gamma l + \sinh \gamma l} \\ &= Z_0 \cdot \frac{\cosh \gamma l \cosh \theta + \sinh \gamma l \sinh \theta}{\cosh \gamma l \sinh \theta + \sinh \gamma l \cosh \theta} \\ &= Z_0 \frac{\cosh (\gamma l + \theta)}{\sinh (\gamma l + \theta)} \\ &= Z_0 \frac{\cosh [(\alpha l + p) + j (\beta l + q)]}{\sinh [(\alpha l + p) + j (\beta l + q)]} \end{aligned}$$

In Chapter 2, page 97, it was seen that :—

$$|\cosh (A + jB)| = \sqrt{\sinh^2 A + \cos^2 B}$$

and $\quad |\sinh (A + jB)| = \sqrt{\sinh^2 A + \sin^2 B}$

Thus
$$|Z_{IN}| = |Z_0| \frac{\sqrt{\sinh^2 (\alpha l + p) + \cos^2 (\beta l + q)}}{\sqrt{\sinh^2 (\alpha l + p) + \sin^2 (\beta l + q)}}$$

i.e., squaring $$|Z_{IN}|^2 = |Z_0|^2 \frac{\sinh^2(\alpha l + p) + \cos^2(\beta l + q)}{\sinh^2(\alpha l + p) + \sin^2(\beta l + q)}$$

Note that $\sin(\beta l + q)$ and $\cos(\beta l + q)$ are periodic functions, whereas $\sinh(\alpha l + p)$ is not.

$|Z_{IN}|^2$ will have a *maximum* value approximately when $\cos^2(\beta l + q) = 1$ and $\sin^2(\beta l + q) = 0$ simultaneously; *i.e.*, when $\beta l + q = \pi, 2\pi, 3\pi$, etc., or, in general, when $\beta l + q = n\pi$, where n is an integer. Similarly $|Z_{IN}|^2$ will have a *minimum* value when $\cos^2(\beta l + q) = 0$ and $\sin^2(\beta l + q) = 1$ simultaneously; *i.e.*, when $\beta l + q = \frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}$, *etc.*, or, in general, when $\beta l + q = (2n - 1)\frac{\pi}{2}$.

Thus $|Z_{IN}|$ will have a maximum value when :—

$$\beta l + q = n\pi$$

i.e., when $$\beta = \frac{n\pi - q}{l} \text{ radians} \tag{45}$$

and the maxima will occur at frequencies for which β takes these values.

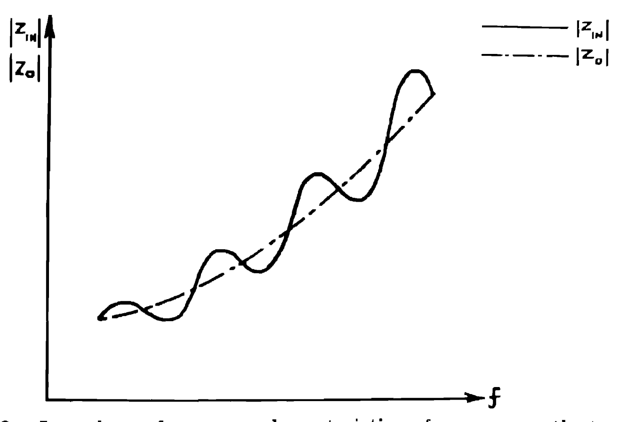


FIG. 756.—Impedance-frequency characteristics of an incorrectly terminated coil-loaded underground cable employing half-section terminations.

Similarly $|Z_{IN}|$ will have a minimum value when :—

$$\beta l + q = (2n - 1)\frac{\pi}{2}$$

i.e., when $$\beta = \frac{(2n - 1)\frac{\pi}{2} - q}{l} \text{ radians} \tag{46}$$

Fig. 756 shows the impedance-frequency characteristic of an incorrectly terminated coil-loaded underground cable employing half-section terminations.

Particular case (i). Line on open-circuit

In this case $Z_R = \infty$, *i.e.*, $\frac{Z_0}{Z_R} = \tanh 0 = 0$

i.e., $\tanh(p + jq) = 0$, *i.e.*, $p = q = 0$

Maxima will occur at frequencies for which :—

$$\beta = \frac{n\pi}{l} \text{ radians} \tag{47}$$

Minima will occur at frequencies for which :—

$$\beta = \frac{(2n-1)\frac{\pi}{2}}{l} \text{ radians} \tag{48}$$

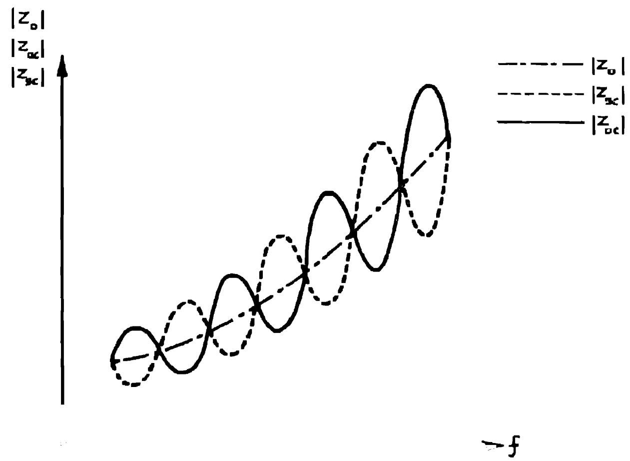


FIG. 757.—Impedance frequency characteristics of a line.

Particular case (ii). Line on short-circuit

In this case $Z_R = 0$ $\therefore \frac{Z_0}{Z_R} = \infty$

Hence $\tanh(p + jq) = \infty$ and $p = 0$, $q = \frac{\pi}{2}$

Maxima occur at frequencies for which :—

$$\beta = \frac{n\pi - \frac{\pi}{2}}{l} = \frac{(2n-1)\frac{\pi}{2}}{l} \text{ radians} \tag{49}$$

Minima occur at frequencies for which :—

$$\beta = \frac{(2n-1)\frac{\pi}{2} - \frac{\pi}{2}}{l} = \frac{(n-1)\frac{\pi}{2}}{l} \text{ radians}$$

or
$$\beta = \frac{n\pi}{l} \text{ radians} \tag{50}$$

since n is any integer.

Comparing equations 47 with 50, and 48 with 49, it will be seen that the frequencies at which maxima occur with the line on open-circuit are the same as those at which minima occur with the line

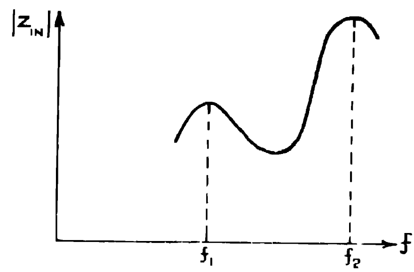


FIG. 758.—Two consecutive maxima on the input impedance-frequency characteristic of a faulty cable.

on short-circuit, and *vice versa* (*see* Fig. 757). In fact, the frequency difference between successive maxima or minima is independent of the value of the terminating impedance.

Fault location using variation of input impedance with frequency

Measurement of variation of input impedance with frequency gives a convenient method of fault location, particularly in the case of loaded underground cables, where trouble may be experienced due to faulty loading coils. An impedance-frequency run is made on the faulty line, and the frequency difference between two successive maxima (or minima) noted (*see* Fig. 758).

Let these frequencies be f_1 and f_2, and let the corresponding values of β at these frequencies be β_1 and β_2. Provided that the frequency difference is small, it may be assumed that the value of q in equation 45 remains unchanged.

At frequency f_1, equation 45 becomes :—

$$\beta_1 = \frac{n\pi - q}{l} \tag{51}$$

At frequency f_2, equation 45 becomes :—

$$\beta_2 = \frac{(n+1)\pi - q}{l} \tag{52}$$

$$\beta_2 - \beta_1 = \frac{\pi}{l}$$

For a loaded underground cable :—

$$\beta = \omega\sqrt{LC}$$

$$\therefore \qquad \omega_2\sqrt{LC} - \omega_1\sqrt{LC} = \frac{\pi}{l}$$

or
$$l = \frac{1}{2\sqrt{LC}\,(f_2 - f_1)} \text{ miles} \tag{53}$$

This gives the distance in miles from the sending end to the point of discontinuity (the fault) and enables the fault to be localised.

EQUIVALENT NETWORKS

When interest lies only in the voltage and current at the two ends of a line, it is convenient to consider the equivalent network.

Equivalent T section

Let the T section shown in Fig. 759*b* represent a line of length l, characteristic impedance Z_0 and propagation constant γ.

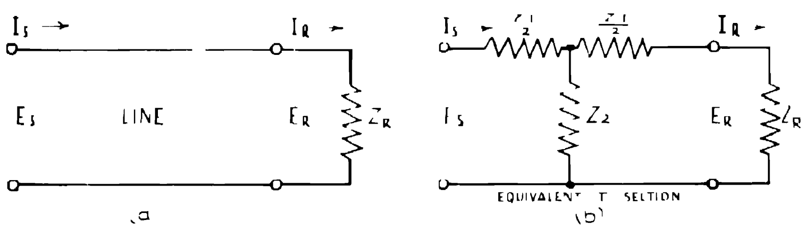


FIG. 759.—Transmission line and its equivalent T section.

Let the sent current and voltage be I_S and E_S, and let the received current and voltage be I_R and E_R.

From equation 17 :—

$$I_R = I_S \cosh \gamma l - \frac{E_S}{Z_0} \sinh \gamma l \tag{54}$$

Considering the T section :—

$$E_S = I_S\frac{Z_1}{2} + (I_S - I_R)\, Z_2$$

$$\therefore \qquad I_R Z_2 = \frac{I_S Z_1}{2} + I_S Z_2 - E_S$$

$$\therefore \qquad I_R = I_S\left(\frac{Z_1}{2Z_2} + 1\right) - \frac{E_S}{Z_2} \tag{55}$$

Equation 55 must be identical with 54 if the T section is to represent the line.

Hence
$$\frac{Z_1}{2Z_2} + 1 = \cosh \gamma l \quad \text{(coefficient of } I_S\text{)} \tag{56}$$

and
$$-\frac{1}{Z_2} = -\frac{\sinh \gamma l}{Z_0} \quad \text{(coefficient of } E_S\text{)} \tag{57}$$

From equation 57 :—

$$Z_2 = \frac{Z_0}{\sinh \gamma l} \tag{58}$$

From equation 56 :—

$$\frac{Z_1}{2} = Z_2 (\cosh \gamma l - 1)$$

$$= Z_0 \frac{\cosh \gamma l - 1}{\sinh \gamma l} \qquad (59)$$

$$= Z_0 \tanh \frac{\gamma l}{2} \qquad (60)$$

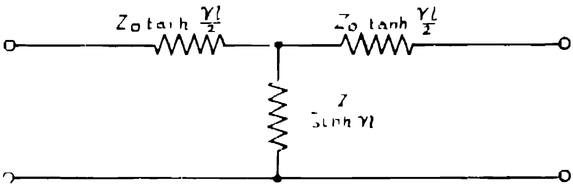


FIG. 760.—Equivalent T section of a uniform transmission line.

Hence the equivalent T section is as shown in Fig. 760.

Equivalent π section

In a similar manner, it can be shown that the equivalent π section is as shown in Fig. 761.

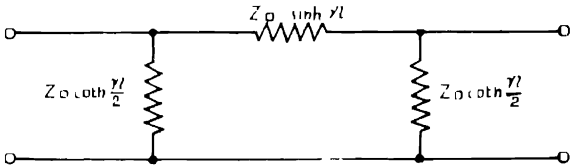


FIG. 761.—Equivalent π section of a uniform transmission line.

Equivalent lattice section

The equivalent lattice section is as shown in Fig. 762.

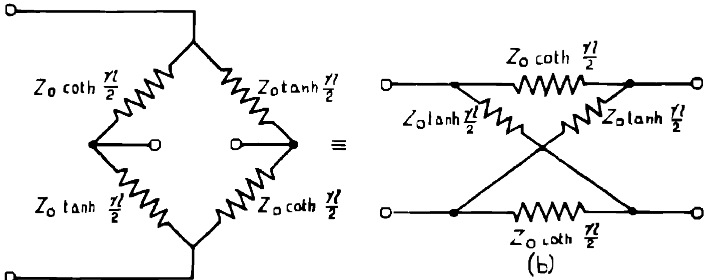


FIG. 762.—Equivalent lattice section of a uniform transmission line.

PROPAGATION CONSTANT OF A LUMP-LOADED LINE

The propagation constant of a lump-loaded line may be determined by considering the equivalent T section.

Let Z_0 be the characteristic impedance of the line before loading ;
Let γ be the propagation constant of the line before loading
Let Z_L be the impedance of the loading coil ;
Let d be the coil spacing ;
Let γ' be the propagation constant of the line after loading.

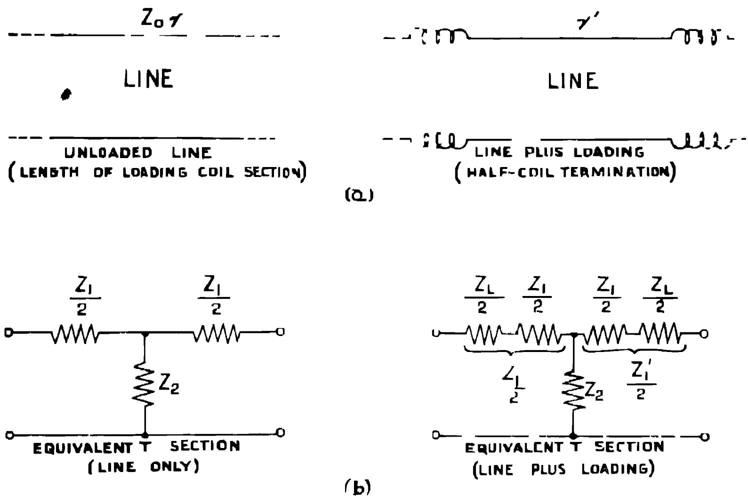


FIG 763 —Equivalent circuits for unloaded and coil-loaded lines

Consider the equivalent T section of a loading coil section (*see* Fig. 763).

Let $$\frac{Z_1}{2} + \frac{Z_L}{2} = \frac{Z_1'}{2}$$

But, from equation 59 :–

$$\frac{Z_1}{2} = Z_0 \frac{\cosh \gamma d - 1}{\sinh \gamma d}$$

Hence $$\frac{Z_1'}{2} = \frac{Z_L}{2} + Z_0 . \frac{\cosh \gamma d - 1}{\sinh \gamma d}$$

From equation 58 :—

$$Z_2 = \frac{Z_0}{\sinh \gamma d}$$

From equation 56 :—

$$\cosh \gamma' d = 1 + \frac{Z_1'}{2Z_2} = 1 + \frac{\sinh \gamma d}{2Z_0}\left[Z_L + 2Z_0 . \frac{\cosh \gamma d - 1}{\sinh \gamma d}\right]$$

$$= 1 + \frac{Z_L \sinh \gamma d}{2Z_0} + \cosh \gamma d - 1$$

i.e.,
$$\cosh \gamma' d = \cosh \gamma d + \frac{Z_L}{2Z_0} \sinh \gamma d \qquad (61)$$

This is known as Campbell's Formula for a loaded line, and it gives an expression for the propagation constant of a loaded section. The method by which this formula was obtained applies only for a loading section with a half-coil termination (*see* Fig. 749*a*). By applying a somewhat similar analysis to a loading section with a half-section termination, an identical expression is obtained. Campbell's Formula therefore applies in both cases.

It can be shown from Campbell's Formula that a loaded line has a low-pass filter characteristic; for on writing $\gamma' = \alpha' + j\beta'$ and replacing γ and Z_0 by their values in terms of the primary line constants (*i.e.*, equations 9 and 11 on pages 751–2), equation 61 becomes:— $\cosh \alpha' d \,.\, \cos \beta' d + j \,.\, \sinh \alpha' d \,.\, \sin \beta' d$

$$= \cosh \left\{ d\sqrt{(R + j\omega L)(G + j\omega C)} \right\}$$
$$+ \frac{Z_L}{2\sqrt{\dfrac{R + j\omega L}{G + j\omega C}}} \cdot \sinh \left\{ d\sqrt{(R + j\omega L)(G + j\omega C)} \right\} \qquad (62)$$

In order to determine the cut-off frequency, it is necessary to neglect the resistance and leakance and to consider a no-loss line; *i.e.*, putting $R = 0$, $G = 0$ and $Z_L = j\omega L_L$ in equation 62 :–

$$\cosh \alpha' d \,.\, \cos \beta' d + j \sinh \alpha' d \,.\, \sin \beta' d = \cos \omega d\sqrt{LC} - \frac{\omega L_L}{2}\sqrt{\frac{C}{L}} \,.\, \sin \omega d\sqrt{LC}$$

Equating real and imaginary parts :—

$$\cosh \alpha' d \,.\, \cos \beta' d = \cos \omega d\sqrt{LC} - \frac{\omega L_L}{2}\sqrt{\frac{C}{L}} \sin \omega d\sqrt{LC} \qquad (63)$$

and $\sinh \alpha' d \,.\, \sin \beta' d = 0$ (64)

These equations have a solution $\alpha' = 0$ if :—

$$\cos \beta' d = \cos \omega d\sqrt{LC} - \frac{\omega L_L}{2}\sqrt{\frac{C}{L}} \sin \omega d\sqrt{LC} \qquad (65)$$

In the practical case $\omega d\sqrt{LC}$ is a small angle (being the phase shift in the loading coil section), so that in equation 65 the approximations :—

$$\sin \omega d\sqrt{LC} = \omega d\sqrt{LC}$$

and
$$\cos \omega d\sqrt{LC} = 1$$

may be made without much loss of accuracy, giving :—

$$\cos \beta' d = 1 - \frac{\omega^2 L_L C d}{2} \qquad (66)$$

This is possible for small values of ω, the limiting value ω_c being given when $\cos \beta' d = 1$

i.e.,
$$\omega_c = \frac{2}{\sqrt{L_L C d}} \tag{67}$$

This means that equations 63 and 64 have a solution $\alpha' = 0$ for all values of ω between $\omega = 0$ and $\omega = \omega_c$; in other words, the loaded line behaves like a low-pass filter, and ω_c gives the cut-off frequency.

Now suppose that the total inductance of the loaded section may be regarded as being uniformly distributed throughout the loading section; let the resultant inductance per mile be denoted by L'; then :—

$$L'd = L_L + Ld \simeq L_L$$

Equation 67 becomes :—

$$\omega_c = \frac{2}{d\sqrt{L'C}} \tag{68}$$

or

$$f_c = \frac{1}{\pi d\sqrt{L'C}} \tag{69}$$

This formula for the cut-off frequency agrees very closely with observed results. In the same way, for calculating the characteristic impedance and the propagation constant of a loaded line, there is very little loss in accuracy if the "smoothing approximation" is made, *i.e.* if the added inductance is considered to be uniformly distributed along the line. The method by which Z_0 and γ are normally calculated is demonstrated in the example on page 722 (Chapter 16), and this method gives perfectly satisfactory results over the normal working frequency range—that is, up to 0·7 of the cut-off frequency.

For frequencies near cut-off, the propagation constant may be computed more accurately by using Campbell's Formula (equation 61); but this is very tedious unless a graphical method (*see* Appendix II) is used for computing the complex hyperbolic functions, and in this case much of the accuracy is lost. For computing accurately the attenuation constant, certain formulae have been derived from Campbell's Formula that enable the attenuation per loading coil section to be calculated simply and accurately. The method given below (without proof) is a modification of a formula obtained by Mayer.

Let R, G, L and C be the primary constants per mile of the unloaded cable. Let L_L be the inductance, R_L the AC resistance, and C_L the shunt capacity of a loading coil, and d the length of the loading coil section in miles.

The following definitions are then made :—

$$L'd = L_L + Ld$$

where L' is the "smoothed" inductance per mile of the loaded cable.

$$C'd = C_L + Cd$$

where C' is the "smoothed" capacity per mile of the loaded cable.

TABLE XXIV

VALUES OF K_1 AND K_2 FOR USE IN EQUATIONS 71, 72 and 73
Propagation Constant of a Lump-loaded Line

x	K_1	K_2	x	K_1	K_2
0.30	0.5172	0.5502	0.65	0.4963	0.6909
0.31	0.5169	0.5523	0.66	0.4959	0.6988
0.32	0.5165	0.5544	0.67	0.4955	0.7072
0.33	0.5156	0.5560	0.68	0.4953	0.7161
0.34	0.5150	0.5581	0.69	0.4953	0.72[illegible]
0.35	0.5144	0.5602	0.70	0.4949	0.7350
0.36	0.5142	0.5626	0.71	0.4949	0.7455
0.37	0.5133	0.5649	0.72	0.4951	0.7565
0.38	0.51[illegible]9	0.5675	0.73	0.4952	0.7681
0.39	0.5123	0.5702	0.74	0.4957	0.7807
0.40	0.5117	0.57[illegible]8	0.75	0.4961	0.7938
0.41	0.5109	0.5754	0.76	0.4968	0.8080
0.42	0.5105	0.5786	0.77	0.4975	0.8227
0.43	0.5100	0.5817	0.78	0.4987	0.8390
0.44	0.5094	0.5849	0.79	0.5000	0.8563
0.45	0.5081	0.5880	0.80	0.5017	0.8752
0.46	0.5077	0.5912	0.81	0.5036	0.8951
0.47	0.5072	0.5948	0.82	0.5060	0.9172
0.48	0.5066	0.5985	0.83	0.5090	0.9413
0.49	0.5058	0.6022	0.84	0.5125	0.9676
0.50	0.5053	0.6064	0.85	0.5172	0.9965
0.51	0.5047	0.6101	0.86	0.5216	1.029
0.52	0.5039	0.6148	0.87	0.5270	1.065
0.53	0.5031	0.6190	0.88	0.5345	1.105
0.54	0.5025	0.6237	0.89	0.5434	1.151
0.55	0.5017	0.6284	0.90	0.5539	1.204
0.56	0.5012	0.6337	0.91	0.5670	1.266
0.57	0.5005	0.6389	0.92	0.5838	1.340
0.58	0.5001	0.6447	0.93	0.6048	1.428
0.59	0.4995	0.6505	0.94	0.6321	1.539
0.60	0.4988	0.6563	0.95	0.6694	1.681
0.61	0.4982	0.6626	0.96	0.7229	1.875
0.62	0.4978	0.6694	0.97	0.8106	2.160
0.63	0.4973	0.6762	0.98	0.9487	2.638
0.64	0.4965	0.6830	0.99	1.289	3.721
—	—	—	1.00	∞	∞

$$f_0 = \frac{1}{\pi d\sqrt{L'C'}}$$

i.e., f_0 is the cut-off frequency calculated in the normal manner using the smoothing approximation.

Let $$Z_d = \sqrt{\frac{L'}{C'}}$$

and $$x = \frac{f}{f_c}\ ,$$ where f is the frequency at which the attenuation is to be determined.

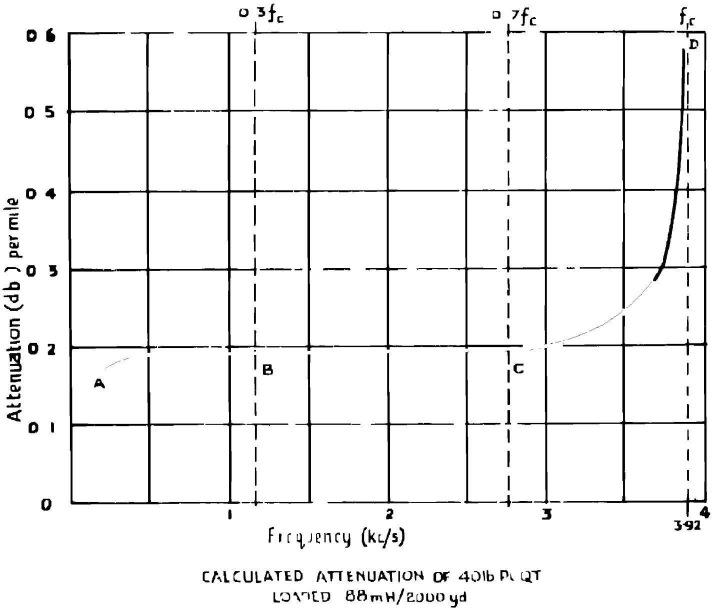


FIG. 764 —Attenuation-frequency characteristic of a 40 lb. PCQT line, loaded 88 mH/2000 yds.

The attenuation αd per loading coil section is then given by :—

$$\alpha d = \alpha_1 + \alpha_2 + \alpha_3 \tag{70}$$

where :—

α_1 is the conductor attenuation per loading coil section, and is given by :—

$$\alpha_1 = \frac{Rd}{Z_d} K_1 \text{ nepers} \tag{71}$$

α_2 is the " coil " attenuation per loading coil section, and is given by :—

$$\alpha_2 = \frac{R_L}{Z_d} . K_1 \text{ nepers} \tag{72}$$

α_3 is the "leakance" attenuation per loading coil section, and is given by :—

$$\alpha_3 = d \quad G \quad Z_d \quad K_2 \text{ nepers} \tag{73}$$

The quantities K_1 and K_2 are dependent on the ratio $x = \frac{f}{f_0}$ of the frequency considered to the cut-off frequency and their values are tabulated in Table XXIV for values of x from 0 3 to 1·0.

Fig 764 shows the attenuation-frequency curve for 40 lb PCQT underground cable loaded 88mH/2,000 yds (the cable specified in the example on page 743) The method of calculating this curve is as follows —

(i) From A to B, calculate α from $\gamma = \sqrt{(R + j\omega L)(G + j\omega C)}$ using the smoothing approximation

(ii) From B to D, calculate α from the method just given.

It will be noted that the flat portion of the curve from B to C is very close to the value obtained in the example on page 743, *i.e*, 0·191 db per mile Thus for frequencies up to 0·7 of the cut-off frequency, there is no advantage to be gained by the more accurate method, and, since a loaded cable is rarely worked at frequencies above this figure, the method of page 742 is the one generally employed

INDEX

VOLUME I

Pages 1 — 875

A

B

D

E

F

G

H

I

M

N

O

P

Q

Page

Page

W

X

Y

Z

(97931) Wt. 3935/9275 K.36 /51 Hw. (TS.22268)